





APPLICANT: Yves R. Laroche, Laurent S. Jespers,  
APPLICANT: Yannick G.J. Ganssemans, Matthew Moyle,  
APPLICANT: Peter W. Bergum  
TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT  
TITLE OF INVENTION: PROTEIN  
NUMBER OF SEQUENCES: 356  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,399  
FILING DATE: June 5, 1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/326,110  
FILING DATE: October 18, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 213/270  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 84 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Ancylostoma caninum  
US-08-486-399-59

Query Match 25.6%; Score 100; DB 2; Length 84;  
Best Local Similarity 30.4%; Pred. No. 0.00071;  
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

QY 11 KCPSENEIFSRCDGR-CQRC-----PNVPRKPLCIRKAPGCVCLGILRNK 57  
DB 5 QCGENEKYDSCGSKCKCKYDGVVEEDDEPNV---PCLVRYCHQDCVCEGFRNMD 61

QY 58 KVCVPRSKC 66  
DB 62 DKCVSAEDC 70

RESULT 5  
US-08-461-965-59  
Sequence 59, Application US/08461965  
Patent No. 5872098  
GENERAL INFORMATION:  
APPLICANT: George P. Vlasuk, Patric H. Stanssens,  
APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,  
APPLICANT: Yves R. Laroche, Laurent S. Jespers,  
APPLICANT: Yannick G.J. Ganssemans, Matthew Moyle,  
APPLICANT: Peter W. Bergum  
TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT  
TITLE OF INVENTION: PROTEIN  
NUMBER OF SEQUENCES: 356  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/461,965  
FILING DATE: June 5, 1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/326,110  
FILING DATE: October 18, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 210/243  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 84 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Ancylostoma caninum  
US-08-461-965-59

Query Match 25.6%; Score 100; DB 2; Length 84;  
Best Local Similarity 30.4%; Pred. No. 0.00071;  
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

QY 11 KCPSENEIFSRCDGR-CQRC-----PNVPRKPLCIRKAPGCVCLGILRNK 57  
DB 5 QCGENEKYDSCGSKCKCKYDGVVEEDDEPNV---PCLVRYCHQDCVCEGFRNMD 61

QY 58 KVCVPRSKC 66  
DB 62 DKCVSAEDC 70

RESULT 6  
US-08-634-641-59  
Sequence 59, Application US/08634641  
Patent No. 5955294  
GENERAL INFORMATION:  
APPLICANT: Vlasuk, George P. Vlasuk  
APPLICANT: Stanssens, Patrick Eric Hugo  
APPLICANT: Mensens, Joris Hilda Llieven  
APPLICANT: Lauwereys, Marc Josef  
APPLICANT: Laroche, Yves Rene  
APPLICANT: Jespers, Laurent Stephane  
APPLICANT: Ganssemans, Yannick Georges Jozef  
APPLICANT: Moyle, Matthew  
APPLICANT: Bergum, Peter W.  
TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT  
TITLE OF INVENTION: PROTEIN  
NUMBER OF SEQUENCES: 356  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071

```

Query Match      25.6%; Score 100; DB 2; Length 84;
Best Local Similarity 30.4%; Pred. No. 0.00071;
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3.
OY      11 KCPSEIFSRGCR-CQRPC-----PNVVPKPLCIRICAPGCVCRLGYLRNK 57
       :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      5 QCGNEKEKDSCGSKCECDKKCKRIDGVVEEDDEEPNV---PCLVRVCHQDCVCEGFRRNKD 61
       |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY      58 KVCVPRSKC 66
       || |
Db      62 DKCVSASEDC 70

RESULT 7
US-09-249-471-59
Sequence 59, Application US/09249471
Patent No. 6040441
GENERAL INFORMATION:
APPLICANT: Vlasuk, George Phillip
APPLICANT: Stanssens, Patrick Eric Hugo
APPLICANT: Messens, Joris Hilda Llieven
APPLICANT: Lauwereys, Marc Josef
APPLICANT: Laroche, Yves Rene
APPLICANT: Jespers, Laurent Stephane
APPLICANT: Ganssemans, Yannick Georges Jozef
APPLICANT: Moyle, Matthew
APPLICANT: Bergum, Peter W.
TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
```

```

; TITLE OF INVENTION:  INHIBITORS AND ANTICOAGULANT
;
; TITLE OF INVENTION:  PROTEIN
;
; NUMBER OF SEQUENCES:  356
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  Lyon & Lyon
; STREET:  633 West Fifth Street
; STREET:  Suite 4700
; CITY:  Los Angeles
; STATE:  California
; COUNTRY:  U.S.A.
;
; ZIP:  90071
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
; MEDIUM TYPE:  storage
;
; COMPUTER:  IBM Compatible
; OPERATING SYSTEM:  IBM P.C. DOS 5.0
; SOFTWARE:  Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/09/249,471
; FILING DATE:
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  08/809,455
; FILING DATE:  April 17, 1997
; APPLICATION NUMBER:  PCT/US95/13231
; FILING DATE:  October 17, 1995
; APPLICATION NUMBER:  08/486,399
; FILING DATE:  June 5, 1995
; APPLICATION NUMBER:  08/486,397
; FILING DATE:  June 5, 1995
; APPLICATION NUMBER:  08/465,380
; FILING DATE:  June 5, 1995
; APPLICATION NUMBER:  08/461,965
; FILING DATE:  June 5, 1995
; APPLICATION NUMBER:  08/326,110
; FILING DATE:  October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME:  BIGGS, SUZANNE L.
; REGISTRATION NUMBER:  30,158
; REFERENCE/DOCKET NUMBER:  216/270
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  (213) 489-1600
; TELEFAX:  (213) 955-0440
; TELEX:  67-3510
;
; INFORMATION FOR SEO ID NO:  59:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  84 amino acids
; TYPE:  amino acid
; TOPOLOGY:  linear
; MOLECULE TYPE:  peptide
; ORIGINAL SOURCE:
; ORGANISM:  Ancylostoma caninum
;
; US-09-249-471-59
;
Query Match                25.6%; Score 100; DB 3; Length 84;
Best Local Similarity      30.4%; Pred. No. 0.00071;
Matches    21; Conservative   8; Mismatches    24; Indels   16; Gaps     3
;
QY      11 KCPSENFESKQDER-CQRC-----PNVVPKPLCIKICAPGCVCRLGLYLRNK 57
       :|::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       5 QCGNERYKSGSKMCKKYDGVVEEDDEPNN---PCLVRVCHODCCVCEEGFYRNKD 61
;
QY      58 KVCYPRSKC 66
       |||_|
Db      62 DKCVSADC 70
;
RESULT      8
US-09-249-472--59
; Sequence 59, Application US/09249472
; Patent No. 6046318
; GENERAL INFORMATION:
; APPLICANT:  Vlasuk, George Phillip
;

```

APPLICANT: Stanssens, Patrick Eric Hugo  
APPLICANT: Messens, Joris Hilda Llieven  
APPLICANT: Lauwereys, Marc Josef  
APPLICANT: Laroche, Yves Rene  
APPLICANT: Jaspers, Laurent Stephane  
APPLICANT: Ganssemans, Yannick Georges Jozef  
APPLICANT: Moyle, Matthew  
APPLICANT: Bergum, Peter W.  
TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE  
TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT  
TITLE OF INVENTION: PROTEIN  
NUMBER OF SEQUENCES: 356  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon 8 Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/249,472  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/809,455  
FILING DATE: April 17, 1997  
APPLICATION NUMBER: PCT/US95/13231  
FILING DATE: October 17, 1995  
APPLICATION NUMBER: 08/486,399  
FILING DATE: June 5, 1995  
APPLICATION NUMBER: 08/486,397  
FILING DATE: June 5, 1995  
APPLICATION NUMBER: 08/465,380  
FILING DATE: June 5, 1995  
APPLICATION NUMBER: 08/461,965  
FILING DATE: June 5, 1995  
APPLICATION NUMBER: 08/326,110  
FILING DATE: October 18, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 216/270  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 84 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Ancylostoma caninum  
US-09-249-472-59

Query Match 25.6%; Score 100; DB 3; Length 84;  
Best Local Similarity 30.4%; Pred. No. 0.00071;  
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

QY 11 KCSNFIHSRCDGR-CORF-----PNVVPRLCIKICAPGCVCHLGLRK 57  
DB 5 QCGENETDCSGKEDCKKCKYGVDEEDDEBNV---PCLVRYCHODCVCEGFRKMD 61  
QY 58 KVCVPRSKC 66  
11 1

DB 62 DKCVSAEDC 70  
RESULT 9  
US-09-249-451-59  
Sequence 59, Application US/09249451  
Patent No. 6087487  
GENERAL INFORMATION:  
APPLICANT: Vlasuk, George Phillip  
APPLICANT: Stanssens, Patrick Eric Hugo  
APPLICANT: Messens, Joris Hilda Llieven  
APPLICANT: Lauwereys, Marc Josef  
APPLICANT: Laroche, Yves Rene  
APPLICANT: Jaspers, Laurent Stephane  
APPLICANT: Ganssemans, Yannick Georges Jozef  
APPLICANT: Moyle, Matthew  
APPLICANT: Bergum, Peter W.  
TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE  
TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT  
TITLE OF INVENTION: PROTEIN  
NUMBER OF SEQUENCES: 356  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon 8 Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/249,451  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/809,455  
FILING DATE: April 17, 1997  
APPLICATION NUMBER: PCT/US95/13231  
FILING DATE: October 17, 1995  
APPLICATION NUMBER: 08/486,399  
FILING DATE: June 5, 1995  
APPLICATION NUMBER: 08/486,397  
FILING DATE: June 5, 1995  
APPLICATION NUMBER: 08/465,380  
FILING DATE: June 5, 1995  
APPLICATION NUMBER: 08/461,965  
FILING DATE: June 5, 1995  
APPLICATION NUMBER: 08/326,110  
FILING DATE: October 18, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 216/270  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 84 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Ancylostoma caninum  
US-09-249-451-59

Query Match 25.6%; Score 100; DB 3; Length 84;

Best Local Similarity 30.4%; Pred. No. 0.00071;  
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

OY 11 KCPSEIFSRDGR-CQRC-----PNNVPRKPLCIKICAPGCVCRGLYLRNK 57  
Db 5 QCGENEKIDSCGSKEDCKCKYDGEVEEDEDEPNV---PCLVAVCHODCVCEGEGFYRND 61  
OY 58 KCVPRSKC 66  
Db 62 DKCVSAEDC 70

## RESULT 10

US-08-809-455-59  
Sequence 59, Application US/08809455  
Patent No. 6090916

## GENERAL INFORMATION:

APPLICANT: Vlasuk, George Phillip  
APPLICANT: Stanssens, Patrick Eric Hugo  
APPLICANT: Messens, Joris Hilda Lieven  
APPLICANT: Lauwereys, Marc Josef  
APPLICANT: Jespers, Laurent Stephane  
APPLICANT: Ganssemans, Yannick Georges Jozef  
APPLICANT: Moyle, Matthew  
TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE  
TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT  
NUMBER OF SEQUENCES: 356  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071

## COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/809,455  
FILING DATE: April 17, 1997

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/13231  
FILING DATE: October 17, 1995  
APPLICATION NUMBER: 08/486,399  
FILING DATE: June 5, 1995  
APPLICATION NUMBER: 08/486,397  
FILING DATE: June 5, 1995  
APPLICATION NUMBER: 08/465,380  
FILING DATE: June 5, 1995  
APPLICATION NUMBER: 08/461,965  
FILING DATE: June 5, 1995  
APPLICATION NUMBER: 08/326,110  
FILING DATE: October 18, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 216/270  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510

## INFORMATION FOR SEQ. ID NO: 59:

SEQUENCE CHARACTERISTICS:  
LENGTH: 84 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: Peptide  
ORIGINAL SOURCE:  
ORGANISM: Ancylostoma caninum  
US-08-809-455-59

Query Match 25.6%; Score 100; DB 3; Length 84;  
Best Local Similarity 30.4%; Pred. No. 0.00071;  
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

OY 11 KCPSEIFSRDGR-CQRC-----PNNVPRKPLCIKICAPGCVCRGLYLRNK 57  
Db 5 QCGENEKIDSCGSKEDCKCKYDGEVEEDEDEPNV---PCLVAVCHODCVCEGEGFYRND 61  
OY 58 KCVPRSKC 66  
Db 62 DKCVSAEDC 70

## RESULT 11

US-09-249-461-59  
Sequence 59, Application US/09249461  
Patent No. 6096877

## GENERAL INFORMATION:

APPLICANT: Vlasuk, George Phillip  
APPLICANT: Stanssens, Patrick Eric Hugo  
APPLICANT: Messens, Joris Hilda Lieven  
APPLICANT: Lauwereys, Marc Josef  
APPLICANT: Jespers, Laurent Stephane  
APPLICANT: Ganssemans, Yannick Georges Jozef  
APPLICANT: Moyle, Matthew  
TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE  
TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT  
NUMBER OF SEQUENCES: 356  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071

## COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/249,461  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/809,455  
FILING DATE: April 17, 1997  
APPLICATION NUMBER: PCT/US95/13231  
FILING DATE: October 17, 1995  
APPLICATION NUMBER: 08/486,399  
FILING DATE: June 5, 1995  
APPLICATION NUMBER: 08/486,397  
FILING DATE: June 5, 1995  
APPLICATION NUMBER: 08/465,380  
FILING DATE: June 5, 1995  
APPLICATION NUMBER: 08/461,965  
FILING DATE: June 5, 1995  
APPLICATION NUMBER: 08/326,110  
FILING DATE: October 18, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 216/270

APPLICATION NUMBER: 08/465,380

NAME: BIGGS, ST

APPLICATION NUMBER: 08/326,110  
FILING DATE: October 18, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.





Thu Jul 21 13:32:05 2001

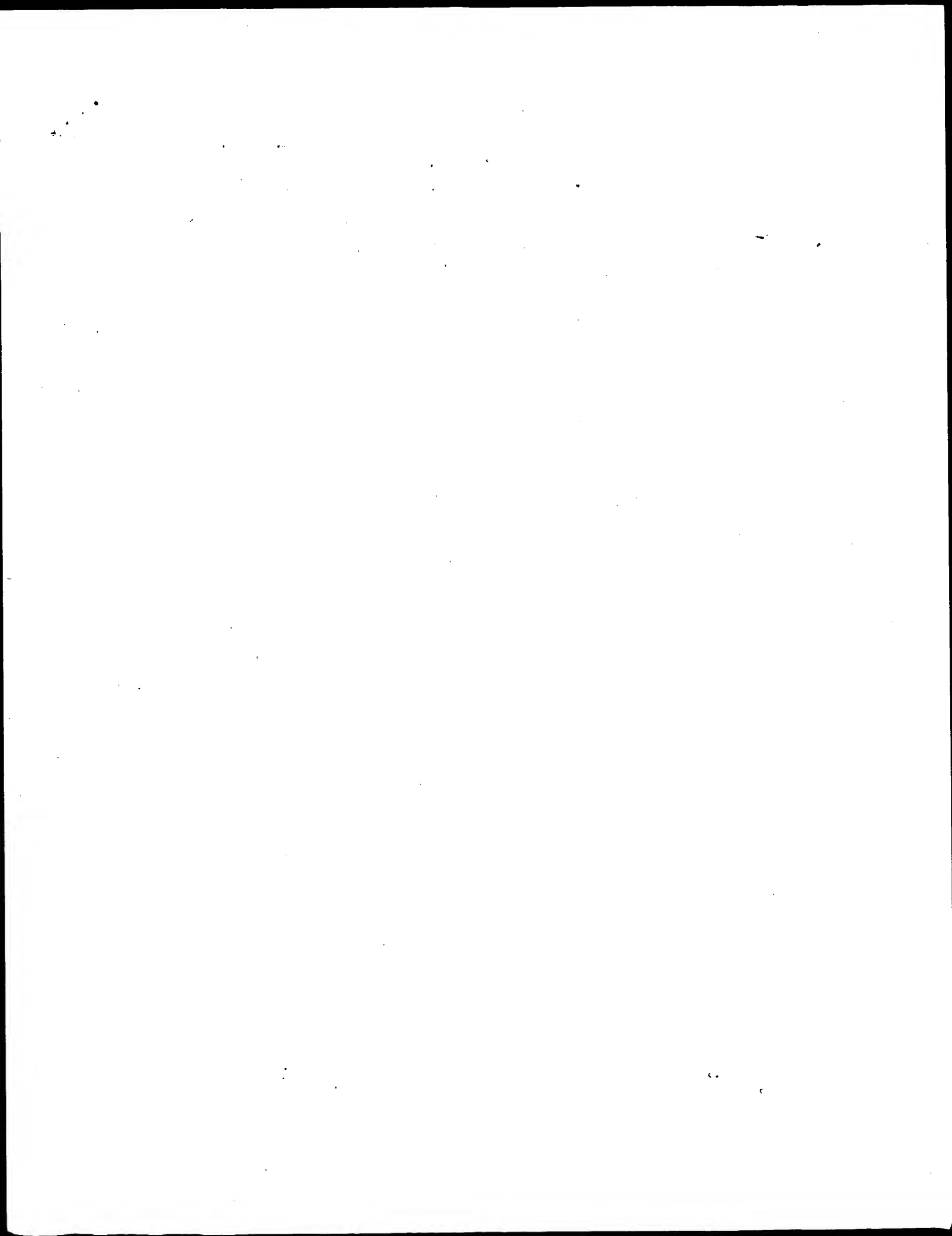
us-09-506-978-1.raii

Page 9

Query Match	25.6%;	Score 100;	DB 2;	Length 91;
Best Local Similarity	30.4%;	Pred. No. 0.00077;		
Matches 21; Conservative	8;	Mismatches 24;	Indels 16;	Gaps 3;

Qy	11	KCSNTEFNSCDGR	-CQRC-----	PNVVRKPLCTKICAGCGCTGRITRANK	57
Db	12	QCGENKHYDSCGSGKECDKCKCTGDEVEEDDEEPNV	---PCLIVKCHQDVCYCEGFTYRND	68	
Qy	58	KVCVPSKSC	66		
Db	69	DKCVSAEDC	77		

Search completed: June 21, 2001, 10:50:43  
Job time: 406 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2001, 10:50:32 ; Search time 9.57 Seconds  
(without alignments)

239.824 Million cell updates/sec

Title: US-09-506-978-1

Sequence: 1 GGFGLGGRGKCPSEIFSR.....CRLGLYLNKKKVCVPSKCG 67

Scoring table: BLASTUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107.5	27.5	2476	1	ZAN_PIG
2	104.5	26.7	56	1	P56682 apis mellif
3	103.5	26.5	2813	1	WVF_CANPA
4	101.5	26.0	937	1	WVF_BOVIN
5	98	25.1	115	1	A6ZF_DROME
6	98	25.1	2813	1	WVF_HUMAN
7	91.5	23.4	2482	1	WVF_PIG
8	90.5	23.1	3133	1	HMC2_BOMMO
9	86	22.0	1700	1	BAR3_CHITE
10	80.5	20.6	63	1	ICE1_ASCSU
11	80.5	20.6	99	1	YOJ2_CAEL
12	80	20.5	1370	1	IGIR_RAT
13	80	20.5	1373	1	IGIR_MOUSE
14	78.5	20.1	65	1	ICE2_ASCSU
15	76.5	19.6	72	1	MT11_MYTED
16	76.5	19.6	902	1	ST14_MOUSE
17	75.5	19.3	349	1	WC22_VARY
18	75.5	19.3	5179	1	MGC2_HUMAN
19	75	19.2	956	1	TSP3_MOUSE
20	75	19.2	4289	1	TENX_HUMAN
21	73	18.7	60	1	MT_PERFL
22	72.5	18.5	1416	2	VN81_CAEL
23	72	18.4	326	1	VT2_MYXYL
24	72	18.4	956	1	TSP3_HUMAN
25	72	18.4	1367	1	IGIR_HUMAN
26	71	18.2	1746	1	TENA_PIG
27	70.5	18.0	62	1	ITR1_ASCSU
28	70.5	18.0	72	1	MT12_MYTED
29	69	17.6	60	1	MTB_ONCMY
30	69	17.6	60	1	MTB_ESOLU
31	69	17.6	77	1	PIIF_BOMMO
32	69	17.6	1339	1	ERB3_RAT
33	68.5	17.5	72	1	MT14_MYTED

34	68	17.4	60	1	MT_PLEPL	P07216 pleuronecte
35	68	17.4	60	1	MT_PSEAM	P55945 pseudopleur
36	68	17.4	4660	1	LRP2_RAT	P98158 rattus norv
37	67.5	17.3	182	1	KROC_SHEEP	P26372 ovis aries
38	67	17.1	62	1	MT4_HUMAN	P47944 homo sapien
39	67	17.1	325	1	YRPI_CAEL	Q10043 caenorhabd
40	67	17.1	2911	1	FBN2_HUMAN	P35556 homo sapien
41	66.5	17.0	171	1	KR2A_SHEEP	P02438 ovis aries
42	66.5	17.0	400	1	PR72_HUMAN	P22891 homo sapien
43	66.5	17.0	1808	1	TENA_CHICK	P10039 gallus gall
44	66	16.9	60	1	MTB_SALSA	P52720 salmo salar
45	65.5	16.8	780	1	DCMA_METHH	O27743 methanobact

## ALIGNMENTS

RESULT	ID	Sequence	Standard	PRT	AA
1	ZAN_PIG	028983	01-NOV-1997 (Rel. 35, Created)		
AC	028983	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	ZONADHESIN PRECURSOR.				
GN	ZAN				
OS	Sus scrofa (Pig).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.				
OX	NCBI_TaxID=9823;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MEISHAN; TISSUE=Testis;				
RA	MEDLINE=96064658; PubMed=7592795;				
RT	Hardy D.M., Garbers D.L.; that binds in a species-specific manner to				
RT	the egg extracellular matrix is homologous to von Willebrand				
RT	factor.";				
RL	J. Biol. Chem. 270:26025-26028(1995).				
CC	-1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA				
CC	OF THE EGG.				
CC	-1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.				
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.				
CC	-1- TISSUE SPECIFICITY: IN TESTIS, PRIMARILY IN HAPLOID SPERMATIDS.				
CC	NOT IN LUNG, LIVER, HEART, SPLEEN, BRAIN, KIDNEY, EPIDIDYMIS.				
CC	-1- DOMAIN: THE UNIQUE N-TERMINAL DOMAIN AND THE MUCIN-LIKE DOMAINS				
CC	ARE MISSING FROM THE ZONADHESIN THAT BINDS TO THE EGG				
CC	EXTRACELLULAR MATRIX. PROCESSING MIGHT OCCUR DURING SPERM				
CC	MATURATION AND/OR CAPACITATION.				
CC	-1- DOMAIN: THE UNIQUE N-TERMINAL DOMAIN PROBABLY MEDIATES SPERM				
CC	ADHESION TO THE ZONA PELLUCIDA.				
CC	-1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,				
CC	THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF				
CC	SPERMATOZOEA OR PROMOTING ADHESION TO THE OVUDUCTAL ISTHMUS.				
CC	-1- DOMAIN: THE WVFED DOMAINS 1 AND 2 MAY MEDIATE COVALENT				
CC	OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).				
CC	-1- SIMILARITY: CONTAINS 1 NAM DOMAIN.				
CC	-1- SIMILARITY: CONTAINS 4.5 WVFED DOMAINS.				
CC	-1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.				
CC	-1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.				
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	EMBL, U40024; AAC48486.1; -				
DR	InterPro: IPR000561; -				
DR	InterPro: IPR000098; -				
DR	InterPro: IPR001846; -				
DR	InterPro: IPR002919; -				

DR Pfam: PF00629; MAM; 2.  
 DR Pfam: PF01826; TIL; 5.  
 DR Pfam: PF00094; vwd; 4.  
 DR PROSITE: PS00022; EGF\_1; 1.  
 DR PROSITE: PS00740; MAM\_1; 1.  
 DR PROSITE: PS01186; EGF\_2; 4.  
 DR PROSITE: PS00060; MAM\_2; 2.  
 KW Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;  
 FT Repeat.  
 FT SIGNAL 1 29  
 FT CHAIN 30 2476  
 FT DOMAIN 30 2418  
 FT TRANSSEM 2419 2439  
 FT DOMAIN 2440 2476  
 FT DOMAIN 144 312  
 FT DOMAIN 319 687  
 FT DOMAIN 688 799  
 FT DOMAIN 800 1184  
 FT DOMAIN 1185 1573  
 FT DOMAIN 1574 1968  
 FT DOMAIN 1969 2370  
 FT DOMAIN 2366 2402  
 FT DISULFID 2370 2381  
 FT DISULFID 2375 2390  
 FT DISULFID 2392 2401  
 FT CARBOHYD 109 109  
 FT CARBOHYD 269 269  
 FT CARBOHYD 735 735  
 FT CARBOHYD 758 758  
 FT CARBOHYD 833 833  
 FT CARBOHYD 1154 1154  
 FT CARBOHYD 1329 1329  
 FT CARBOHYD 1448 1448  
 FT CARBOHYD 1544 1544  
 FT CARBOHYD 1556 1556  
 FT CARBOHYD 1654 1654  
 FT CARBOHYD 1843 1843  
 FT CARBOHYD 1965 1965  
 FT CARBOHYD 2122 2122  
 FT CARBOHYD 2165 2165  
 FT CARBOHYD 2178 2178  
 FT CARBOHYD 2329 2329  
 FT CARBOHYD 2359 2359  
 SO SEQUENCE 2476 AA; 270364 MW; A13B690373A6548C CRC64;

Query Match 27.5%; Score 107.5; DB 1; Length 2476;  
 Best Local Similarity 33.3%; Pred. No. 0.00079;  
 Matches 22; Conservative 9; Mismatches 18; Indels 17; Gaps 3;

OY 11 KCPSEIERSRCRCGRC-----ORCPNVVPRPLCTIKICAPGCVRLGRLNKKKVCV 61  
 DB 1455 KCPSSSYSTCANPCPATCLSLNPSYCPSTLP-----CAEGEGCKHGLISGTS-CV 1506

OY 62 PRSKCG 67  
 DB 1507 PLUSCG 1512

RESULT 2  
 AMCI\_APIME STANDARD; PRT; 56 AA.  
 ID AMCI\_APIME  
 AC P56682;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE CHYMOTRYPSIN INHIBITOR (AMCI).  
 OS Apis mellifera (Honeybee).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;  
 OC Apoidea; Apidae; Apis.  
 NCBI\_TaxID=7460;

RN [1]  
 RP SEQUENCE, AND STRUCTURE BY NMR.  
 RC TISSUE=Hemolymph;  
 RX MEDLINE=9933935; PubMed=10411628;  
 RA Bania J., Stachowiak D., Polanowski A.;  
 RT "Primary structure and properties of the cathepsin G/chymotrypsin  
 RT inhibitor from the larval hemolymph of Apis mellifera.";  
 RL Eur. J. Biochem. 262:680-687(1999).  
 CC - FUNCTION: CHYMOTRYPSIN AND CATHEPSIN G INHIBITOR.  
 CC - SUBCELLULAR LOCATION: SECRETED.  
 DR PDB: 1CCV; 12-MAR-99.  
 DR InterPro: IPR002919; .  
 DR Pfam: PF01826; TIL; 5.  
 KW Serine protease inhibitor; 3D-structure.  
 FT DISULFID 3 36  
 FT DISULFID 12 32  
 FT DISULFID 16 28  
 FT DISULFID 20 56  
 FT DISULFID 38 50  
 SO SEQUENCE 56 AA; 5973 MW; 092B2815A6EB2B7F CRC64;

Query Match 26.7%; Score 104.5; DB 1; Length 56;  
 Best Local Similarity 36.8%; Pred. No. 8.9e-05;  
 Matches 21; Conservative 7; Mismatches 26; Indels 3; Gaps 2;

OY 11 KCPSEIERSRCRCGRCRCRVNVRKP-LCTIKICAPGCVRLGRLNKKKVCVPRSKC 66  
 DB 2 ECGFNEVNTGCSACAPTCAG--PKTRICTMOCRCRIGCCGCGFLRNEGACVLPENC 56

RESULT 3  
 VWF\_CANFA STANDARD; PRT; 2813 AA.  
 ID VWF\_CANFA  
 AC Q28295; Q28311; Q9T514;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE VON WILLEBRAND FACTOR PRECURSOR (VWF).  
 GN F8VWF OR VWF.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Stoy S.J., Shibuya H., Noneman D.J., Holzhauer J., Mohammed I.H.,  
 RA Johnson G.S.;  
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RA Montgomery R.R., Fahs S., Montgomery M.W.;  
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Venta P.J., Li J., Yuzbasian-Gurkan V., Brewer G.J., Schall W.D.;  
 RT "Complete sequence of the structural gene for canine von Willebrand  
 RT factor and identification of a mutation causing Scottish terrier von  
 RT Willebrand's disease.";  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 1234-1669 FROM N.A.  
 RC TISSUE=Blood;  
 RA Mancuso D.J., Christopherson P.A., Kroner P.A., Montgomery R.R.;  
 RT "The canine von Willebrand factor gene: sequence and expression of  
 RT a region encoding the glycoprotein ID/IX binding domain.";  
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.  
 CC - FUNCTION: IMPORTANT IN THE MAINTENANCE OF HEMOSTASIS, IT  
 CC PARTICIPATES IN PLATELET-VESSEL WALL INTERACTIONS BY FORMING A  
 CC NONCOVALENT COMPLEX WITH COAGULATION FACTOR VIII AT THE SITE OF  
 CC VASCULAR INJURY (BY SIMILARITY).  
 CC - SUBUNIT: MULTIMERIC (BY SIMILARITY).  
 CC - TISSUE SPECIFICITY: BLOOD.

CC -1- P.M. ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR  
 CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 3 VWEA DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 3 VMFC DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 4 VMFD DOMAINS.  
 CC -1- SIMILARITY: CONTAINS A C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).  
 CC -1- SIMILARITY: SOME, TO SILKORM HEMOCYTIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: I76227; AAB05549.1; -  
 DR EMBL: AF099154; AAD04919.1; -  
 DR EMBL: U66246; AAB93766.1; -  
 DR HSSP: P04275; IAT2.  
 DR InterPro: IPR000359; -  
 DR InterPro: IPR001007; -  
 DR InterPro: IPR001846; -  
 DR InterPro: IPR002035; -  
 DR InterPro: IPR002919; -  
 DR Pfam: PF00007; Cys\_Knot; 1.  
 DR Pfam: PF01826; TIL; 4.  
 DR Pfam: PF00092; VWA; 3.  
 DR Pfam: PF00093; VWC; 3.  
 DR Pfam: PF00094; VWD; 4.  
 DR PRINTS: PR00453; VWEADOMAIN.  
 DR PROSITE: PS01185; CTCK\_1; 1.  
 DR PROSITE: PS01225; CTCK\_2; 1.  
 DR PROSITE: PS0234; VWEA; 3.  
 DR PROSITE: PS01208; VMFC; 3.  
 DR Blood coagulation: Platelet: Glycoprotein: Extracellular matrix;  
 KW Plasma; Endothelial cell; Repeat; Cell adhesion; Signal.  
 FT SIGNAL 1 22  
 FT PROPEP 23 763  
 FT CHAIN 764 2813  
 FT DOMAIN 35 179  
 FT DOMAIN 388 541  
 FT DOMAIN 764 787  
 FT DOMAIN 788 853  
 FT DOMAIN 826 853  
 FT DOMAIN 867 1013  
 FT DOMAIN 1277 1453  
 FT DOMAIN 1498 1665  
 FT DOMAIN 1691 1871  
 FT DOMAIN 1950 2102  
 FT DOMAIN 2216 2261  
 FT DOMAIN 2255 2326  
 FT DOMAIN 2429 2495  
 FT DOMAIN 2580 2650  
 FT DOMAIN 2724 2812  
 FT SITE 531 533  
 FT SITE 698 700  
 FT SITE 2507 2509  
 FT DISULFID 767 808  
 FT DISULFID 776 804  
 FT DISULFID 810 821  
 FT DISULFID 867 996  
 FT DISULFID 889 1031  
 FT DISULFID 898 993  
 FT DISULFID 914 921  
 FT DISULFID 1060 1084  
 FT DISULFID 1071 1111  
 FT DISULFID 1089 1091  
 FT DISULFID 1153 1165  
 FT DISULFID 1149 1169  
 FT DISULFID 1126 1130  
 FT DISULFID 1196 1199

FT DISULFID 1234 1237 BY SIMILARITY.  
 FT DISULFID 1272 1458 BY SIMILARITY.  
 FT DISULFID 1669 1670 BY SIMILARITY.  
 FT DISULFID 1686 1872 BY SIMILARITY.  
 FT DISULFID 1879 1904 BY SIMILARITY.  
 FT DISULFID 1899 1940 OR 1942 (BY SIMILARITY).  
 FT DISULFID 1972 2123 BY SIMILARITY.  
 FT DISULFID 1950 2085 BY SIMILARITY.  
 FT DISULFID 1927 2088 BY SIMILARITY.  
 FT DISULFID 1993 2001 BY SIMILARITY.  
 FT DISULFID 2724 2774 BY SIMILARITY.  
 FT DISULFID 2739 2788 BY SIMILARITY.  
 FT DISULFID 2750 2804 BY SIMILARITY.  
 FT DISULFID 2754 2806 BY SIMILARITY.  
 FT CARBOHYD 99 2811 BY SIMILARITY.  
 FT CARBOHYD 156 156 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 211 211 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 666 666 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 857 857 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1231 1231 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1515 1515 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1574 1574 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2223 2223 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2290 2290 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2357 2357 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2400 2400 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2546 2546 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2585 2585 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2790 2790 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 55 55 G -> E (IN REF. 2).  
 FT CARBOHYD 70 70 V -> I (IN REF. 3).  
 FT CARBOHYD 266 266 A -> G (IN REF. 2).  
 FT CARBOHYD 280 280 I -> V (IN REF. 2).  
 FT CARBOHYD 409 411 VCH -> ICO (IN REF. 2).  
 FT CARBOHYD 994 994 G -> A (IN REF. 1).  
 FT CARBOHYD 1021 1021 F -> L (IN REF. 2).  
 FT CARBOHYD 1021 1021 L -> P (IN REF. 2).  
 FT CARBOHYD 2381 2381 P -> L (IN REF. 2).  
 FT CARBOHYD 2406 2406 P -> L (IN REF. 2).  
 FT CARBOHYD 2813 2813 AA: 309716 MW: 5DF93E1E5E72E80C CRC64;  
 SO SEQUENCE

Query Match 26.5%; Score 103.5; DB 1; Length 2813;  
 Best Local Similarity 28.8%; Pred. No. 0.0023;  
 Matches 19; Conservative 15; Mismatches 27; Indels 5; Gaps 2;

OY 2 GFGGAGRGKCPNSNEIFSRCDGRCQFPCPNV-VPRPLKIKICAPGCVRLGYLRKKKVC 60  
 DB 646 GFCAL-----SCFGQGYVLDGCGTPCNMTGSLSPBEDCNEVCLGCGFCPPGLYLDERGDC 701  
 OY 61 VPRSKC 66  
 DB 702 VPRKQC 707

RESULT 4  
 ID VWE BOVIN STANDARD; PRT; 937 AA.  
 AC P80012; Q28011;  
 DT 01-NOV-1995 (rel. 32, Created)  
 DT 01-OCT-2000 (rel. 40, Last sequence update)  
 DT 01-OCT-2000 (rel. 40, Last annotation update)  
 DE VON WILLEBRAND FACTOR PRECURSOR (VWF) (FRAGMENT).  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 RX MDLINE=97307751; PubMed=9155093;  
 RA Janel N., Ribba A.S., Chereil G., Kerbiriou-Nabias D., Meyer D.;

[illegible]



RA Lynch D.C., Zimmerman T.S., Collins C.J., Brown M., Morin M.J.,  
 RA Ling E.H., Livingston D.M.;  
 RT "Molecular cloning of cDNA for human von Willebrand factor:  
 RT authentication by a new method.";  
 RL Cell 41:49-56(1985).  
 RN [14]  
 RP REVISIONS.  
 RA Lynch D.C.;  
 RL Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.  
 RN [15]  
 RP SEQUENCE OF 2731-2813 FROM N.A.  
 RX MEDLINE-87260814; PubMed-3496594;  
 RA Collins C.J., Underdahl J.P., Levene R.B., Ravera C.P.,  
 RA Morin M.J., Dombalagian M.J., Ricca G., Livingston D.M.,  
 RA Lynch D.C.;  
 RT "Molecular cloning of the human gene for von Willebrand factor and  
 RT identification of the transcription initiation site.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4393-4397(1987).  
 RN [16]  
 RP DISULFIDE BONDS.  
 RX MEDLINE-88163465; PubMed-3502076;  
 RA Marti T., Rossetti S.J., Titani K., Walsh K.A.;  
 RT "Identification of disulfide-bridged substructures within human von  
 RT Willebrand factor";  
 RL Biochemistry 26:8099-8109(1987).  
 RN [17]  
 RP STRUCTURE OF CARBOHYDRATES.  
 RX MEDLINE-86274702; PubMed-3089784;  
 RA Samor B., Michalski J.C., Debrey H., Mazurier C., Goudemand M.,  
 RA van Halbeek H., Vliegenhart J.F.G., Montreuil J.;  
 RT "Primary structure of a new tetraantennary glycan of the N-  
 RT acetylglucosaminic type isolated from human factor VIII/von  
 RT Willebrand factor.";  
 RL Eur. J. Biochem. 158:295-298(1986).  
 RN [18]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 1261-1468.  
 RX MEDLINE-98221174; PubMed-9553097;  
 RA Emsley J., Cruz M., Handin R., Liddington R.;  
 RT "Crystal structure of the von Willebrand factor A1 domain and  
 RT implications for the binding of platelet glycoprotein Ib.";  
 RL J. Biol. Chem. 273:10396-10401(1998).  
 RN [19]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 1685-1873.  
 RX MEDLINE-97472999; PubMed-9331419;  
 RA Huizinga E.G., Martin van der Plas R., Kroon J., Sixma J.J., Gros P.;  
 RT "Crystal structure of the A3 domain of human von Willebrand factor:  
 RT implications for collagen binding.";  
 RL Structure 5:1147-1156(1997).  
 RN [20]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1686-1872.  
 RX MEDLINE-97460108; PubMed-9312128;  
 RA Bielewska J., Cruz M., Atieno A., Handin R., Liddington R.;  
 RT "The von Willebrand factor A3 domain does not contain a metal ion-  
 RT dependent adhesion site motif.";  
 RL J. Biol. Chem. 272:25162-25167(1997).  
 RN [21]  
 RP VARIANTS TRP-1597 AND ASP-1607.  
 RX MEDLINE-89264495; PubMed-2786201;  
 RA Ginsburg D., Konkle B.A., Gill J.C., Montgomery R.R.,  
 RA Bockstaele P.L., Johnson T.A., Yang A.Y.;  
 RT "Molecular basis of human von Willebrand disease: analysis of  
 RT platelet von Willebrand factor mRNA.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:3723-3727(1989).  
 RN [22]  
 RP VARIANT TRP-1628.  
 RX MEDLINE-91196734; PubMed-1673047;  
 RA Iannuzzi M.C., Hidak N., Boehnke M., Bruck M.E., Hanna W.T.,  
 RA Collins F.S., Ginsburg D.;  
 RT "Analysis of the relationship of von Willebrand disease (vWD) and  
 RT hereditary hemorrhagic telangiectasia and identification of a  
 RL potential type IIA vWD mutation (I1685 to Thr).";  
 RL Am. J. Hum. Genet. 48:757-763(1991).  
 RN [23]

RP VARIANTS NORMANDY-2 AND NORMANDY-3.  
 RX MEDLINE-92001464; PubMed-1832934;  
 RA Gaucher C., Mercier B., Jorjeux S., Oufkir D., Mazurier C.;  
 RT "Identification of two point mutations in the von Willebrand factor  
 RT gene of three families with the 'Normandy' variant of von Willebrand  
 RT disease.";  
 RL Br. J. Haematol. 78:506-514(1991).  
 RN [24]  
 RP VARIANT CYS-1308.  
 RX MEDLINE-92104315; PubMed-1761120;  
 RA Donner M., Andersson A.-C., Nilsson I.M.,  
 RA Dahlback B., Holmberg L.;  
 RT "An Arg545-->Cys545 substitution mutation of the von Willebrand  
 RT factor in type IIB von Willebrand's disease.";  
 RL Eur. J. Haematol. 47:342-345(1991).  
 RN [25]  
 RP VARIANTS TRP-1306; CYS-1308 AND PRO-1613.  
 RX MEDLINE-91185601; PubMed-2010538;  
 RA Randi A.M., Rabinowitz I., Mancuso D.J., Mannucci P.M., Sadler J.E.;  
 RT "Molecular basis of von Willebrand disease type IIB. Candidate  
 RT mutations cluster in one disulfide loop between proposed platelet  
 RT glycoprotein Ib binding sequences.";  
 RL J. Clin. Invest. 87:1220-1226(1991).  
 RN [26]  
 RP VARIANTS TRP-1306; CYS-1308; MET-1316; GLN-1341 AND HIS-1399.  
 RX MEDLINE-91185602; PubMed-1672694;

Query Match 25.1%; Score 98; DB 1; Length 2813;  
 Best Local Similarity 26.0%; Pred. No. 0.0085;  
 Matches 19; Conservative 13; Mismatches 27; Indels 14; Gaps 2;

OY 8 GRG-----KCPSEIFRCDCRCORCPNV-VKPKLCIKCAGCCYRGL 53  
 DB 635 GCGVAVARREPCRCCLNCPKCGVYLQCTPCNLTCRSLSPDECNCLEGCFCPPGLY 694  
 OY 54 RNRKRVCPRSKC 66  
 DB 695 MDERGDCVPRKAC 707

RESULT 7  
 VWF-PIG STANDARD; PRT; 2482 AA.  
 AC Q28833;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE VON WILLEBRAND FACTOR PRECURSOR (VWF) (FRAGMENT).  
 GN FBWVF OR VWF.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Seaman W.T., Read M.S., Bellinger D.A., Nichols T.C.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 397-553 FROM N.A.  
 RX MEDLINE-93356762; PubMed-8352759;  
 RA Laveigne J.H., Piao Y.C., Ferreira V., Kerblirou-Nabias D.,  
 RA Bahak B.R., Meyer D.;  
 RT "Primary structure of the factor VIII binding domain of human, porcine  
 RT and rabbit von Willebrand factor";  
 RL Biochem. Biophys. Res. Commun. 194:1019-1024(1993).  
 CC -!- FUNCTION: IMPORTANT IN THE MAINTENANCE OF HOMEOSTASIS, IT  
 CC PARTICIPATES IN PLATELET-VESSEL WALL INTERACTIONS BY FORMING A  
 CC NONCOVALENT COMPLEX WITH COAGULATION FACTOR VIII AT THE SITE OF  
 CC VASCULAR INJURY (BY SIMILARITY).  
 CC -!- SUBUNIT: MULTIMERIC (BY SIMILARITY).  
 CC -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR  
 CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).  
 CC -!- SIMILARITY: CONTAINS 3 VWF-C DOMAINS.



-1- SIMILARITY: CONTAINS A C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).  
 CC -1- SIMILARITY: SOME, TO SILKMOB HEMOCYTIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/announce/or\\_send\\_an\\_email\\_to\\_license@sib-sib.ch](http://www.isb-sib.ch/announce/or_send_an_email_to_license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AF052036; AAC06229.1; -  
 DR EMBL: S64541; AAB27829.2; -  
 DR HSSP: P04275; IATF. -  
 DR InterPro: IPR000359; -  
 DR InterPro: IPR001007; -  
 DR InterPro: IPR001846; -  
 DR InterPro: IPR002035; -  
 DR InterPro: IPR002919; -  
 DR Pfam: PF00007; Cys\_knotc; 1.  
 DR Pfam: PF00092; vwa; 3.  
 DR Pfam: PF00093; vwc; 3.  
 DR Pfam: PF00094; vwd; 3.  
 DR Pfam: PF01826; TIL; 3.  
 DR PRINTS: PR00365; ENDOTHELIN.  
 DR PRINTS: PR00453; VWFADOMAIN.  
 DR PROSITE: PS0234; VWF; 3.  
 DR PROSITE: PS01208; VWF; 3.  
 DR PROSITE: PS01185; CTCK\_1; 1.  
 DR PROSITE: PS01225; CTCK\_2; 1.  
 DR Blood coagulation; Platelet; Glycoprotein; Extracellular matrix;  
 KW Plasma; Endothelial cell; Repeat; Cell adhesion.  
 FT NON\_TER 1  
 FT PROPEP <1 437  
 FT CHAIN 438 2482  
 FT DOMAIN 62 215  
 FT DOMAIN 438 461  
 FT DOMAIN 462 507  
 FT DOMAIN 500 527  
 FT DOMAIN 541 687  
 FT DOMAIN 947 1127  
 FT DOMAIN 1167 1334  
 FT DOMAIN 1360 1540  
 FT DOMAIN 1619 1771  
 FT DOMAIN 1885 1930  
 FT DOMAIN 1924 1997  
 FT DOMAIN 2098 2164  
 FT DOMAIN 2249 2319  
 FT DOMAIN 2393 2481  
 FT SITE 2176 2178  
 FT DISULFID 441 482  
 FT DISULFID 450 478  
 FT DISULFID 484 495  
 FT DISULFID 541 670  
 FT DISULFID 563 705  
 FT DISULFID 572 667  
 FT DISULFID 588 595  
 FT DISULFID 734 758  
 FT DISULFID 745 785  
 FT DISULFID 763 765  
 FT DISULFID 827 839  
 FT DISULFID 823 843  
 FT DISULFID 800 804  
 FT DISULFID 870 873  
 FT DISULFID 908 911  
 FT DISULFID 942 1128  
 FT DISULFID 1338 1339  
 FT DISULFID 1355 1541  
 FT DISULFID 1348 1573  
 FT DISULFID 1568 1609  
 FT DISULFID 1641 1792  
 FT DISULFID 1619 1754  
 FT DISULFID 1596 1757

FT DISULFID 1662 1670 BY SIMILARITY.  
 FT DISULFID 2393 2443 BY SIMILARITY.  
 FT DISULFID 2408 2457 BY SIMILARITY.  
 FT DISULFID 2419 2473 BY SIMILARITY.  
 FT DISULFID 2423 2475 BY SIMILARITY.  
 FT DISULFID 2423 2480 BY SIMILARITY.  
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 905 905 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1184 1184 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1243 1243 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1640 1640 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1892 1892 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1959 1959 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2026 2026 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2069 2069 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2215 2215 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2254 2254 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2459 2459 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 2482 AA; 272394 MW; D499B/DDFBBCAEDD CRC64;  
 Query Match 23.4%; Score 91.5; DB 1; Length 2482;  
 Best Local Similarity 28.8%; Pred. No. 0.037;  
 Matches 19; Conservative 11; Mismatches 31; Indels 5; Gaps 2;  
 Oy 2 GFGGLGGRGKPSNEIFRCRGRORPCPNV-VKPLCIKICACGCYCRGLNKKKVC 60  
 Db 320 GFCAI-----SCPFGQVYLQCTPCNLTCRSLSYDEBCAEDCLCCPCPGLYDGSQDC 375  
 Oy 61 VPRSKC 66  
 Db 376 VPKAC 381  
 RESULT 8  
 HMCT\_BOMMO  
 ID HMCT\_BOMMO STANDARD; PRT; 3133 AA.  
 AC P96092.  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE HEMOCYTIN PRECURSOR (HUMORAL LECTIN).  
 OS Bombyx mori (Silk moth).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
 OC Bombycoidea; Bombycidae; Bombyx.  
 OX NCBI\_TaxID=7091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-FUYOU X TOKAI; TISSUE=hemocyte;  
 RX MEDLINE=95178544; PubMed=7873598;  
 RA Kotani E., Yamakawa M., Iwamoto S.-I., Tashiro M., Mori H., Sumida M.,  
 RA Matsubara F., Tanaka K., Kadoh-Okuda K., Kato Y., Mori H.;  
 RT "Cloning and expression of the gene of hemocytin, an insect humoral  
 RT lectin which is homologous with the mammalian von Willebrand  
 RT factor.";  
 RL Biochim. Biophys. Acta 1260:245-258(1995).  
 RN [2]  
 RP SEQUENCE OF 2221-3133 FROM N.A.  
 RA Kotani E., Iwamoto S.-I., Tashiro M., Mori H., Sumida M.,  
 RA Matsubara F., Yamakawa M.;  
 RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: ADHESIVE PROTEIN AND RELATES TO HEMOSTASIS OR  
 CC ENCAPSULATION OF FOREIGN SUBSTANCES FOR SELF-DEFENSE.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN HEMOCYTES DURING LARVAL-PUPAL  
 CC METAMORPHOSIS.  
 CC -1- INDUCTION: HEMAGGLUTINATION ACTIVITY IS INCREASED BY BACTERIAL  
 CC OR VIRAL INFECTION AND INHIBITED BY D-MANNOSE, N-ACTYL-D-  
 CC GALACTOSAMINE AND D-MALTOSE.  
 CC -1- PTM: MAY BE CONVERTED INTO THE 260 KDA MATURE HEMOCYTIN BY  
 CC PROTEOLYSIS.  
 CC -1- SIMILARITY: TO MAMMALIAN VON WILLEBRAND FACTOR.  
 CC -1- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT

```

FT CARBOHYD 2276 2276 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2451 2451 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2647 2647 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2647 2647 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2654 2654 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2653 2653 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2794 2794 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2810 2810 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2865 2865 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2929 2929 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2964 2964 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3028 3028 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3028 3028 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 1288 1288 R -> G.
FT VARIANT 1305 1305 T -> S.
SQ SEQUENCE 3133 AA; 343350 MM; E5210D5D14A7B2B2 CRC64;

Query Match 23.1%; Score 90.5; DB 1; Length 3133;
Best Local Similarity 27.5%; Pred. No. 0.056;
Matches 19; Conservative 11; Mismatches 26; Indels 13; Gaps 3;

Qy 11 KCPNSHFRKCDGRQRCRPN-----VYKPLCT-KICAPGV-----CRLGTLRKKK 57
Db 769 KCPPEVYQACAYKCDRLCDHFKKTLIAKGRGISCMYCDVDSVASNGEGSSRWMD 828
Qy 58 KYCVPRSKC 66
Db 829 RTCVPRKDC 837

RESULT 9
BAR3_CHITE
ID BAR3_CHITE STANDARD; PRT; 1700 AA.
AC 003376;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE BALBIANI RING PROTEIN 3 PRECURSOR.
DE BR3.
OS Chironomus tentans (Midge).
CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
CC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
CC Chironomidae; Chironomidae; Chironominae; Chironomus.
OX NCBI_TaxID=7153;
RL [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Salivary gland;
RX MEDLINE=90172404; Pubmed=1689777;
RA Paulsson G., Lendahl U., Gall J., Ericsson C., Wieslander L.;
RT "The Balbiani ring 3 gene in Chironomus tentans has a diverged
RT repetitive structure split by many introns.";
RL J. Mol. Biol. 211:331-349(1990).
CC -1- FUNCTION: USED BY THE LARVAE TO CONSTRUCT A SUPRAMOLECULAR
CC STRUCTURE, THE LARVAL TUBE. BALBIANI RING PROTEIN 3 COULD PLAY A
CC INTRACELLULARLY AND IN THE GLAND LUMEN IN ORDER TO PREVENT THESE
CC FROM FORMING WATER-INSOLUBLE FIBERS TOO EARLY.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: SALIVARY GLAND.
CC -1- DOMAIN: HAS 82 APPROXIMATE REPEATS OF CYS-X-CYS-X-CYS.
CC
CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X52263; CA36506.1; -
DR PIR; S08167; S08167.
DR HSP; P18055; ZMR.
DR InterPro; IPRO00853; -
DR PRINTS; PR00876; MTNEMATODE.

```

KW Repeat: Signal.  
 FT SIGNAL 1 ? POTENTIAL.  
 FT CHAIN ? 1700 BALBICANT RING PROTEIN 3.  
 SQ SEQUENCE 1700 AA; 186145 MW; 34202B28521B0815 CRC64;

Query Match 22.0%; Score 86; DB 1; Length 1700;  
 Best Local Similarity 29.1%; Pred. No. 0.1;  
 Matches 23; Conservative 9; Mismatches 21; Indels 26; Gaps 6;

QY 12 CPSENEFSCDGRGCRFCNNVVKP-----LCIKICAPGCV---CRLGYLRNK 56  
 DB 1234 CPENQIWM--CDNTRCVCCKNMEKPADNCKTKWMNDKCCVCCKPCEGCK--GVNKN 1290  
 QY 57 KKVC---VPRSK-----CG 67  
 DB 1291 ANTCSCCPADKAKKAPASC 1309

RESULT 10  
 ICBL ASCSU STANDARD; PRT; 63 AA.  
 AC P07851;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE CHYMOTRYPSIN/ELASTASE ISOINHIBITOR 1 (C/E-1 INHIBITOR).  
 OS Ascaris suum (Pig roundworm) (Ascaris lumbricoides).  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;  
 OC Ascarididae; Ascaris.  
 OX NCBI\_TaxId=6253;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=84255715; PubMed=6564898;  
 RA Babin D.R., Peanasky R.J., Goos S.M.;  
 RT "The isoforms of chymotrypsin/elastase from Ascaris  
 lumbricoides: the primary structure."  
 RT Arch. Biochem. Biophys. 232:143-161(1984).  
 RL [2]  
 RN X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF COMPLEX WITH ELASTASE.  
 RP MEDLINE=95006335; PubMed=7922044;  
 RX Huang K., Strynadka N.C., Bernard V.D., Peanasky R.J., James M.N.;  
 RT "The molecular structure of the complex of Ascaris  
 chymotrypsin/elastase inhibitor with porcine elastase."  
 RT Structure 2:679-689(1994).  
 RL CC -1- FUNCTION: DEFEND THE ORGANISM AGAINST THE HOST'S PROTEINASES.  
 CC -1- SIMILARITY: BELONGS TO THE ASCARIS TRYPSIN INHIBITOR FAMILY.  
 DR PIR: S07127; S07127.  
 DR PDB: 1EAI; 05-APR-99.  
 DR InterPro: IPR002919; -.  
 DR Pfam: PF01826; TIL; 1.  
 KW Serine protease inhibitor; 3D-structure.  
 FT DISULFID 5 38  
 FT DISULFID 17 29  
 FT DISULFID 14 23  
 FT DISULFID 21 60  
 FT DISULFID 40 54  
 FT ACT SITE 31 32  
 FT ACT SITE 31 32  
 SQ SEQUENCE 63 AA; 6862 MW; 5DC10DE75B375F16 CRC64;

Query Match 20.6%; Score 80.5; DB 1; Length 63;  
 Best Local Similarity 30.3%; Pred. No. 0.031;  
 Matches 20; Conservative 10; Mismatches 23; Indels 13; Gaps 4;

QY 8 GRGCPSENEFSCDGRGCRFCNNVVKP-----LCIKICAPGCV---CRLGYLRNK 60  
 DB 1 GQSGGNEVWTECTGCKMCKGPDENPPCLMKRPSK--ECSFG-----RGMRTYDKGC 54  
 QY 61 VPRSK 66  
 DB 55 IPASOC 60

RESULT 11  
 Y0J2\_CAEEL STANDARD; PRT; 99 AA.  
 ID Y0J2\_CAEEL  
 AC P34625;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-JUN-1994 (Rel. 29, Last annotation update)  
 DE HYPOTHETICAL 10.6 KDA PROTEIN ZK353.2 IN CHROMOSOME III.  
 GN ZK353.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Rhabditinae; Caenorhabditis.  
 OX NCBI\_TaxId=6239;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,  
 RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,  
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
 RA Wohldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans".  
 RL Nature 368:32-38(1994).  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC  
 DR EMBL: L15313; AAA28199.1; .  
 DR WormPep: ZK353.2; CE00386.  
 KW Hypothetical protein.  
 SQ SEQUENCE 99 AA; 10561 MW; 862C659838E47E5F CRC64;

Query Match 20.6%; Score 80.5; DB 1; Length 99;  
 Best Local Similarity 28.1%; Pred. No. 0.043;  
 Matches 18; Conservative 10; Mismatches 17; Indels 19; Gaps 2;

QY 1 GGGGIGGGRGCRPSNEFSCDGRGCRFCNNVVKP-----LCIKICAPGCV---CRLGYLRNK 60  
 DB 31 GGGGIGGGRGCRGADNVFR--WRCCDYSPECCIOLEFWVFLVPIIIGFVCIACGL 88  
 QY 44 PCGV 47  
 DB 89 AGCV 92

RESULT 12  
 IGIR\_RAT STANDARD; PRT; 1370 AA.  
 ID IGIR\_RAT  
 AC P24062;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE INSULIN-LIKE GROWTH FACTOR I RECEPTOR PRECURSOR (EC 2.7.1.112).  
 GN IGIR.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RA MEDLINE=95277910; PubMed=7758167;  
 RX Du J., Delafontaine P.;  
 RT "Inhibition of vascular smooth muscle cell growth through antisense  
 transcription of a rat insulin-like growth factor I receptor cDNA.";  
 RL Cite. Res. 76:963-972(1995).  
 RN [2]  
 RP SEQUENCE OF 1-364 FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY;  
 RA MEDLINE=90017496; PubMed=2477843;  
 RX Werner H., Moloschak M., Adamo M., Shen-Orr Z., Roberts C.T. Jr.,  
 RA Leroith D.;  
 RT "Developmental regulation of the rat insulin-like growth factor I  
 receptor gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:7451-7455(1989).  
 RN [3]  
 RP SEQUENCE OF 913-1017 FROM N.A.  
 RC MEDLINE=92412145; PubMed=1530648;  
 RA Kurauchi H., Jobo K., Ohta M., Kawasaki T., Itoh N.;  
 RT "A new member of the insulin receptor family, insulin  
 receptor-related receptor, is expressed preferentially in the  
 kidney.";  
 RL Biochem. Biophys. Res. Commun. 187:934-939(1992).  
 CC -1- FUNCTION: THIS RECEPTOR BINDS INSULIN-LIKE GROWTH FACTOR I (IGF I)  
 WITH A HIGH AFFINITY AND IGF II WITH A LOWER AFFINITY. IT HAS A  
 TYROSINE-PROTEIN KINASE ACTIVITY.  
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
 TYROSINE-PROTEIN KINASE ACTIVITY.  
 CC -1- SUBUNIT: TETRAMER OF 2 ALPHA AND 2 BETA CHAINS LINKED BY DISULFIDE  
 BONDS. THE ALPHA CHAINS CONTRIBUTE TO THE FORMATION OF THE LIGAND-  
 BINDING DOMAIN, WHILE THE BETA CHAIN CARRIES THE KINASE DOMAIN.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-  
 PROTEIN KINASES.  
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement. (See <http://www.isb-sdb.ch/announce/>  
 or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC  
 CC EMBL: L29232; AAA41392.1; -  
 DR EMBL: M27293; AAA41384.1; -  
 DR PIR: A38837; A38837.  
 DR HSSP: P06213; 11RK.  
 DR InterPro: IPR000494; -  
 DR InterPro: IPR000719; -  
 DR InterPro: IPR001245; -  
 DR InterPro: IPR001777; -  
 DR InterPro: IPR002011; -  
 DR InterPro: IPR002174; -  
 DR Pfam: PF00757; Furlin-like; 1.  
 DR Pfam: PF01030; Recep\_L-domain; 2.  
 DR Pfam: PF00041; fn3; 2.  
 DR Pfam: PF00069; kinase; 1.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR PROSITE: PS00107; PROTEIN KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN KINASE\_TYR; 1.  
 DR PROSITE: PS00239; RECEPTOR\_TYR\_KIN\_1; 1.  
 DR PROSITE: PS00011; PROTEIN KINASE\_DOM; 1.  
 KW Transferrase; Tyrosine-protein kinase; Receptor; Transmembrane;  
 KM Glycoprotein; ATP-binding; Phosphorylation; Repeat; Signal.  
 FT SIGNAL 1 30  
 FT CHAIN 31 741  
 FT CHAIN 742 1370  
 FT CHAIN 742 936  
 DOMAIN 742 936  
 EXTRA-CELLULAR (POTENTIAL).

FT TRANSMEM 937 960  
 FT DOMAIN 961 1370  
 FT DOMAIN 608 829  
 FT DOMAIN 830 929  
 FT DOMAIN 1000 1275  
 FT DOMAIN 1006 1014  
 FT BINDING 1034 1034  
 FT ACT\_SITE 1136 1136  
 FT MOD\_RES 1166 1166  
 FT CARBOHYD 51 51  
 FT CARBOHYD 102 102  
 FT CARBOHYD 135 135  
 FT CARBOHYD 245 245  
 FT CARBOHYD 314 314  
 FT CARBOHYD 418 418  
 FT CARBOHYD 439 439  
 FT CARBOHYD 535 535  
 FT CARBOHYD 608 608  
 FT CARBOHYD 623 623  
 FT CARBOHYD 641 641  
 FT CARBOHYD 748 748  
 FT CARBOHYD 757 757  
 FT CARBOHYD 765 765  
 FT CARBOHYD 901 901  
 FT CARBOHYD 914 914  
 FT CONFLICT 985 986  
 SQ SEQUENCE 1370 AA; 155395 MW; A59468974A1CB145 CRC64;  
 Query Match 20.5%; Score 80; DB 1; Length 1370;  
 Best Local Similarity 33.8%; Pred. No. 0.37;  
 Matches 23; Conservative 5; Mismatches 18; Indels 22; Gaps 4;  
 QY 14 SNEIFSRG--DRCRCFCNVVVKPLCI--KICAGGV-----CRGLGLR 54  
 DB 198 NNEYNRTCTTTRCKQMKPSCVCGKACHTENNCCPCLGSDPDDTTVCACHHY- 256  
 QY 55 NKKKVCVP 62  
 DB 257 --KGVCVP 262  
 RESULT 13  
 ICLR\_MOUSE  
 ID ICLR\_MOUSE STANDARD; PRT; 1373 AA.  
 AC 060751; 062123; 070438;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE INSULIN-LIKE GROWTH FACTOR I RECEPTOR PRECURSOR (EC 2.7.1.112).  
 GN IGF1R.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Navarero M., Garandel V., Barenton B., Bernardi H.;  
 RT "Cloning of cDNA for the mouse insulin-like growth factor I  
 receptor.";  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1-329 FROM N.A.  
 RC STRAIN-CD-1; TISSUE-Kidney;  
 RA Jun W., Liu Z., Alvares K., Kumar A., Wallner E.I., Kanwar Y.S.;  
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1134-1203 FROM N.A.  
 RA MEDLINE=90152381; PubMed=2482828;  
 RX Wilks A.F., Kurban R.R., Hovens C.M., Ralph S.J.;  
 RT "The application of the polymerase chain reaction to cloning members  
 of the protein tyrosine kinase family.";  
 RL Gene 85:67-74(1989).

CC	-1	FUNCTION: THIS RECEPTOR BINDS INSULIN-LIKE GROWTH FACTOR I (IGF 1) WITH A HIGH AFFINITY AND IGF II WITH A LOWER AFFINITY. IT HAS A TYROSINE-PROTEIN KINASE ACTIVITY	
CC	-1	CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP + PROTEIN TYROSINE PHOSPHATE.	
CC	-1	SUBUNIT: TETRAMER OF 2 ALPHA AND 2 BETA CHAINS LINKED BY DISULFIDE BONDS. THE ALPHA CHAINS CONTRIBUTE TO THE FORMATION OF THE LIGAND-BINDING DOMAIN, WHILE THE BETA CHAIN CARRIES THE KINASE DOMAIN.	
CC	-1	SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN	
CC	-1	SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-PROTEIN KINASES.	
CC	-1	SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.	
CC	-1	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement/orsend">http://www.isb-sib.ch/announcement/orsend</a> an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
DR	EMBL:	AF056187: AAC12782.1; -	
DR	EMBL:	U00182: AAC52123.1; -	
DR	EMBL:	M33422: AAA40013.1; -	
DR	HSSP:	P11362: IGF1r.	
DR	MGD:	MG1:96433; IGF1r.	
DR	InterPro:	IPR000494; -	
DR	InterPro:	IPR000719; -	
DR	InterPro:	IPR001245; -	
DR	InterPro:	IPR001777; -	
DR	InterPro:	IPR002011; -	
DR	InterPro:	IPR002174; -	
DR	Pfam:	PF00041; fn3; 2.	
DR	Pfam:	PF00069; PKinase; 1.	
DR	Pfam:	PF00757; Furlin-like; 1.	
DR	Pfam:	PF01030; Recep.L.domain; 2.	
DR	PRINTS:	PR00109; TYRKINASE.	
DR	PROSITE:	PS00107; PROTEIN_KINASE_ATP; 1.	
DR	PROSITE:	PS00109; PROTEIN_KINASE_TYR; 1.	
DR	PROSITE:	PS50011; PROTEIN_KINASE_DOM; 1.	
DR	PROSITE:	PS00239; RECEPTOR_TYR_KIN_1; 1.	
KM	Transferrase:	Tyrosine-protein kinase; Receptor; Transmembrane; Glycoprotein; App-binding; Phosphorylation; Repeat; Signal.	
FT	SIGNAL	1..30	POTENTIAL.
FT	CHAIN	31..741	INSULIN-LIKE GROWTH FACTOR I RECEPTOR, ALPHA-CHAIN..
FT	CHAIN	742..1373	INSULIN-LIKE GROWTH FACTOR I RECEPTOR, BETA-CHAIN..
FT	DOMAIN	742..936	EXTRACELLULAR (POTENTIAL).
FT	TRANSEM	937..960	POTENTIAL.
FT	DOMAIN	961..1373	CYTOPLASMIC (BY SIMILARITY).
FT	DOMAIN	608..829	FIBRONECTIN TYPE-III.
FT	DOMAIN	830..929	FIBRONECTIN TYPE-III.
FT	DOMAIN	1000..1276	PROTEIN KINASE.
FT	NP_BIND	1006..1014	ATP (BY SIMILARITY).
FT	BLINDING	1034..1034	ATP (BY SIMILARITY).
FT	ACT_SITE	1137..1137	BY SIMILARITY.
FT	MOD_RES	1167..1167	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	51..51	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	102..102	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	135..135	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	245..245	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	314..314	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	418..418	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	439..439	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	535..535	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	608..608	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	623..623	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	641..641	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	748..748	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	757..757	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	765..765	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	901..901	N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CARBOHYD	914	914	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CONFLICT	58	59	FL-> LV (IN REF. 2).
FT	CONFLICT	260	260	C-> S (IN REF. 2).
FT	CONFLICT	301	301	D-> G (IN REF. 2).
FT	CONFLICT	306	306	E-> V (IN REF. 2).
FT	CONFLICT	324	324	C-> S (IN REF. 2).
FT	CONFLICT	1134	1134	V-> I (IN REF. 3).
FT	CONFLICT	1145	1145	V-> D (IN REF. 3).
FT	CONFLICT	1202	1202	V-> I (IN REF. 3).
SQ	SEQUENCE	1373	155787	MM; 5E3B72E101B379 CRC64;

	Query Match	20.5%;	Score 80;	DB 1;	Length 1373;
	Best Local Similarity	33.8%;	Pred. No. 0.37;		
	Matches	23;	Conservative	5;	Mismatches 18; Indels 22; Gaps
OY	14 SNEIFSR--DGRORPCPNVPRKPLT-----KITAPCV-----CRLLXLR	54			
Dd	198 NNEYRYRWTTNNRQCKMCPSCGKRACETENNECHPECLSGCHPDDNTTCVCAGRHYT-	256			
OY	55 NKKKYCVP 62				
Dd	257 --KGVCVP 262				

RESULT	14
ICE2_ASCSU	
ID	ICE2_ASCSU
STANDARD;	
PRT;	65 AA

DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE CHYMOTRYPSIN/ELASTASE ISOINHIBITORS 2 TO 5.  
OS ASCARIS suum (Plg roundworm) (Ascaris lumbricoidea).  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea.  
OC Ascarididae; Ascaris.  
OX NCBI\_TaxID=6253;

RP SEQUENCE.  
RX MEDLINE=84255715; PubMed=6564898;  
RA Babin D.R., Peanasky R.J., Goos S.M.;  
RT "The isonibitofors of chymotrypsin/elastase from Ascaris  
RT lumbricoles: the primary structure.";  
RL Arch. Biochem. Biophys. 232:143-161(1984).  
CC -1- FUNCTION: DEFEND THE ORGANISM AGAINST THE HOST'S PROTEINASES  
CC -1- SIMILARITY: BELONGS TO THE ASCARIS TRYPSIN INHIBITOR FAMILY.

DR	Hsp90, PI3K86, IAKD.				
DR	Interpro; IPRO02919; -.				
KW	Pfam; PF01826; TIL; 1.				
KM	Serine protease inhibitor.				
FT	DISULFID	4	37		BY SIMILARITY.
FT	DISULFID	13	32		BY SIMILARITY.
FT	DISULFID	16	28		BY SIMILARITY.
FT	DISULFID	20	59		BY SIMILARITY.
FT	DISULFID	39	53		BY SIMILARITY.
FT	ACT_SITE	31	32		REACTIVE BOND.
FT	VARIANT	25	25		R -> N (IN INHIBITOR 2 AND 4)
FT	VARIANT	40	40		T -> S (IN INHIBITOR 2 AND 4)
FT	VARIANT	64	65		MISSING (IN INHIBITOR 2).
FT	VARIANT	65	65		K -> E (IN INHIBITOR 3).
FT	VARIANT	65	65		K -> R.
FO	SEQUENCE	65 AA:	7241 MW;		BAE5ICA16EAA4BE3 CRC64:

Query Match	20.1%;	Score 78.5;	DB 1;	Length 65;
Best Local Similarity	31.7%;	Pred. No. 0.051;		
Matches 19; Conservative	6;	Mismatches 32;	Indels 3;	Gaps 2

[illegible]

```

RESULT 15
MTLL_MTTED
ID MTLL_MTTED STANDARD; PRT; 72 AA.
AC P80246;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE METALLOTHIONEIN 10-I (MT-10-I).
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=6550;
RN [1]
RP SEQUENCE.
RX MEDLINE=94062828; PubMed=8243463;
RA Mackay E.A., Overnell J., Dunbar B., Davidson I., Hunziker P.E.,
RA Kaegi J.H.R., Fothergill J.E.;
RT "Complete amino acid sequences of five dimeric and four monomeric
RT forms of metallothionein from the edible mussel Mytilus edulis.";
RL Eur. J. Biochem. 218:183-194(1993).
CC -!- FUNCTION: THE METALLOTHIONEINS ARE INVOLVED IN THE CELLULAR
CC SEQUESTRATION OF TOXIC METAL IONS.
CC -!- SUBUNIT: MONOMER.
CC -!- INDUCTION: BY CADMIUM.
CC -!- SIMILARITY: BELONGS TO FAMILY 2 IN METALLOTHIONEIN SUPERFAMILY.
DR PIR: S39416; S39416.
DR HSSP: P15358; 1SK2.
DR InterPro: IPR001008; -.
DR InterPro: IPR003019; -.
DR Pfam: PF00131; metalthio; 1.
DR PRINTS: PR00875; MTMOULUSC.
KW Metal-binding; Metal-thiolate cluster; Chelation; Cadmium.
SQ SEQUENCE 72 AA; 7105 MW; 3EA99D959AEB3B2 CRC64;

```

```

Query Match 19.6%; Score 76.5; DB 1; Length 72;
Best Local Similarity 37.2%; Pred. No. 0.089;
Matches 16; Conservative 5; Mismatches 19; Indels 3; Gaps 2;

```

```

QY 7 GGRGKPSNFIIFSRDGRQRFQGNVVPKPLCTKIKICAPGCVCR 49
DB 33 GADCKGSCGKVCVCKSGRCE--CGKGTGPSTCK-CAFGCSCK 72

```

Search completed: June 21, 2001, 10:52:57  
Job time: 145 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2001, 10:47:32 ; Search time 14.8 seconds  
(without alignments)  
344.844 Million cell updates/sec

Title: US-09-506-978-1

Perfect score: 391  
Sequence: 1 GGFGLGGRGKCPSENEIFSR.....CRLGYLNKKKVCPSKSCG 67

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	133	34.0	245	2 T21654	hypothetical prote
2	118	30.2	561	2 T27318	hypothetical prote
3	118	30.2	626	2 T27319	hypothetical prote
4	113	28.9	2155	2 T30197	alpha tectorin - m
5	113	28.9	5376	2 T42215	zonadhesin - mouse
6	109.5	28.0	802	2 T24293	hypothetical prote
7	109.5	28.0	949	2 T24294	hypothetical prote
8	107.5	27.5	2476	2 T34022	zonadhesin - pig
9	100.5	25.7	137	2 T15609	hypothetical prote
10	100	25.6	140	2 T16574	hypothetical prote
11	98	25.1	2813	1 VWHU	von Willebrand fac
12	97.5	24.9	2120	2 T30243	alpha tectorin - c
13	96.5	24.7	1036	2 T17405	scavenger receptor
14	94.5	24.2	1023	2 T30257	IgE Fc binding pro
15	93	23.8	192	2 T25513	hypothetical prote
16	92.5	23.7	195	2 T28803	hypothetical prote
17	92	23.5	249	2 T24604	hypothetical prote
18	91	23.3	1373	2 JEO095	gastric mucin MUC5
19	90.5	23.1	135	2 T15610	hypothetical prote
20	90.5	23.1	3133	2 S52093	hemocytin - silkw
21	89.5	22.9	1642	2 T19130	hypothetical prote
22	88.5	22.6	145	2 T15608	hypothetical prote
23	86	22.0	1700	2 S08167	Balbiani ring 3 pr
24	84.5	21.6	780	2 A34102	von Willebrand fac
25	84.5	21.6	1513	2 A54895	mucin 2, intestinal
26	83.5	21.4	13288	2 T03099	mucin, submaxillar
27	80.5	20.6	63	2 S07127	chymotrypsin/elast
28	80.5	20.6	99	2 S44658	ZK53.2 protein -
29	80	20.5	1371	2 A33837	insulin-like growt

30	79.5	20.3	100	2 T23389	hypothetical prote
31	79.5	20.3	1506	2 T30886	integumentary muc
32	78	19.9	490	2 T32003	hypothetical prote
33	77.5	19.8	63	2 S08572	chymotrypsin/elast
34	77.5	19.8	915	2 T21773	hypothetical prote
35	77.5	19.8	927	2 T21772	hypothetical prote
36	77	19.7	169	2 T15611	hypothetical prote
37	76.5	19.6	72	2 S39416	hypothetical prote
38	76.5	19.6	2195	2 T34264	metallothionein 10
39	75.5	19.3	253	2 T25768	hypothetical prote
40	75.5	19.3	348	2 T28623	hypothetical prote
41	75.5	19.3	349	2 D72175	G2R protein - vari
42	75.5	19.3	349	2 D36858	gene G4R protein -
43	75.5	19.3	2910	2 T42214	ocogelin - mouse
44	75.5	19.3	3020	2 A43932	mucin 2 precursor,
45	75	19.2	956	1 A46016	thrombospondin 3 ~

## ALIGNMENTS

```

RESULT 1
T21654
hypothetical protein F32D8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T21654
R:Wilkinson, J.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19454
A:Accession: T21654
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-245 <WIL>
A:Cross-references: EMBL:Z74031; PIDN:CAA98455.1; GSPDB:GN00023; CESP:F32D8.3
A:Experimental source: clone F32D8
C:Genetics:
A:Gene: CESP:F32D8.3
A:Map position: 5
A:Introns: 61/1; 83/3
C:Superfamily: Caenorhabditis elegans hypothetical protein F32D8.3

Query Match 34.0% Score 133; DB 2; Length 245;
Best Local Similarity 43.1% Pred. NO. 3.4e-06;
Matches 28; Conservative 7; Mismatches 20; Indels 10; Gaps 4;

QY 6 LGGRGKCPSENEIFSRCDRCORFQPNVVPKPLCIKICAPGCVCRGLYLNKKVCV----61
Db 57 INGEENCPMFQSFHSC--ACESTCNN--PDYCSK-CEPGCTCRNGFVRNSLKCVLPEE 111
QY 62 -PRSK 65
Db 112 CPRK 116

RESULT 2
T27318
hypothetical protein Y69H2.3a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
R:McMurray, A.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z20343
A:Accession: T27318
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-561 <WIL>
A:Cross-references: EMBL:Z98877; PIDN:CAE54472.1; GSPDB:GN00023; CESP:Y69H2.3a
A:Experimental source: clone Y69H2
C:Genetics:
A:Gene: CESP:Y69H2.3a

```

A:Map position: 5  
A:Introns: 183/1; 247/1; 288/1; 318/1; 364/1; 388/2; 457/1; 481/2; 521/3

Query Match 30.2%; Score 118; DB 2; Length 561;  
Best Local Similarity 36.8%; Pred. No. 0.00019;  
Matches 21; Conservative 11; Mismatches 23; Indels 2; Gaps 2;

QY 12 CPSENEIFSRCDGRC-QRFPCPNV-VPRKPLCIKICAPGCVCLGTLRNKKKVCYPRSKC 66  
DB 123 CPVNEVSENECHNPCTERKCPQKNAPQVNCIMACQVCGSCMDGFVRNNGVCVKEAEC 179

RESULT 3  
T27319  
hypothetical protein Y69H2.3b - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T27319  
R:McMurray, A.  
submitted to the EMBL Data Library, August 1997

A:Reference number: T27319  
A:Accession: T20343  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1626 <MIL>

A:Cross-references: EMBL:Z98877; PIDN:CAB54473.1; GSPDB:GN00023; CESP:Y69H2.3b

A:Experimental source: clone Y69H2

C:Genetics:

A:Gene: CESP:Y69H2.3b

A:Map position: 5

A:Introns: 183/1; 247/1; 312/1; 353/1; 383/1; 429/1; 453/2; 522/1; 546/2; 586/3

Query Match 30.2%; Score 118; DB 2; Length 626;  
Best Local Similarity 36.8%; Pred. No. 0.0002;  
Matches 21; Conservative 11; Mismatches 23; Indels 2; Gaps 2;

QY 12 CPSENEIFSRCDGRC-QRFPCPNV-VPRKPLCIKICAPGCVCLGTLRNKKKVCYPRSKC 66  
DB 123 CPVNEVSENECHNPCTERKCPQKNAPQVNCIMACQVCGSCMDGFVRNNGVCVKEAEC 179

RESULT 4  
T30197  
alpha tectorin - mouse

C:Species: Mus musculus (house mouse)

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 21-Jul-2000

C:Accession: T30197

R:Legan, P.K.; Rau, A.; Keene, J.N.; Richardson, G.P.

J. Biol. Chem. 272, 8791-8801, 1997

A:Title: The mouse tectorins. Modular matrix proteins of the inner ear homologous to com

A:Reference number: T20771; MUID:97236843

A:Accession: T30197

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2155 <LEG>

A:Cross-references: EMBL:X99805; NID:91915908; PIDN:CAA6138.1; PID:91915909

A:Experimental source: strain CD1; whole cochleae

A:Note: non-collagenous protein only expressed in the inner ear, by cells both in and su

Query Match 28.9%; Score 113; DB 2; Length 2155;  
Best Local Similarity 38.6%; Pred. No. 0.0016;  
Matches 22; Conservative 7; Mismatches 26; Indels 2; Gaps 2;

QY 11 KCPSENEIFSRCDGRCQRFPCPNVVPKPLCIKICAPGCVCLGTLRNKKKVCYPRSKC 67  
DB 983 ECPENSHREEC-MTCTETCETLALGPICVDSGSCGCDDBGY-RLQSGQCVTRSECG 1037

RESULT 5  
T42215

zonadhesin - mouse  
N:Alternate names: sperm-specific membrane protein  
C:Species: Mus musculus (house mouse)  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T42215  
R:Guo, Z.; Garbers, D.L.  
J. Biol. Chem. 273, 3415-3421, 1998  
A:Title: Species diversity in the structure of zonadhesin, a sperm-specific membrane  
A:Reference number: Z22080; MUID:98123114  
A:Accession: T42215  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-5376 <GAO>  
A:Cross-references: EMBL:U97068; NID:93327420; PID:93327421; PIDN:AMC26680.1  
C:Genetics:  
A:Gene: Zan  
A:Map position: 5  
A:Function:  
A:Description: functions in multiple cell adhesion processes  
A:Note: found exclusively on the apical region of the sperm head  
C:Keywords: cell adhesion

Query Match 28.9%; Score 113; DB 2; Length 5376;  
Best Local Similarity 37.5%; Pred. No. 0.0031;  
Matches 24; Conservative 8; Mismatches 22; Indels 10; Gaps 3;

QY 11 KCPSENEIFSRCDGRCQRFPCPNVVPKPLCIKICAPGCVCLGTLRNKKKVCYPR 63  
DB 4743 KCPANSLYHCLPTCLPSCSN--PDGREGTSHKAPSTCRGCVQDPGLLN-KDTCVHK 4799

QY 64 SKCG 67  
DB 4800 NOCG 4803

RESULT 6  
T24293  
hypothetical protein T01D3.3a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T24293

R:Steward, C.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19870

A:Accession: T24293

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-802 <MIL>

A:Cross-references: EMBL:Z81110; PIDN:CAB03259.1; GSPDB:GN00023; CESP:T01D3.3a

A:Experimental source: clone T01D3

C:Genetics:

A:Gene: CESP:T01D3.3a

A:Map position: 5

A:Introns: 74/1; 121/1; 200/2; 493/1; 673/3; 772/2

Query Match 28.0%; Score 109.5; DB 2; Length 802;  
Best Local Similarity 39.0%; Pred. No. 0.0017;  
Matches 23; Conservative 7; Mismatches 24; Indels 5; Gaps 3;

QY 12 CPSENEIFSRCDGRCQRFPCPNVVPKPLCIKICAPGCVCLGTLRNKKKVCYPRSKC 66  
DB 82 CGANEQYSAFCSSQPCOD-BSTPACPAFCQPCICLPGYIRDSPPRSACVPRGIC 139

RESULT 7  
T24294  
hypothetical protein T01D3.3b - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T24294  
R:Steward, C.



submitted to the EMBL Data Library, October 1996  
 A:Reference number: Z19870  
 A:Accession: T24294  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA

A:Residues: 1-949 <WILL>  
 A:Cross-references: EMBL:Z81110; PIDN:CA803260.1; GSPDB:GN00023; CESP:T01D3.3b

A:Experimental source: clone T01D3  
 C:Genetics:

A:Gene: CESP:T01D3.3b  
 A:Map position: 5

A:Introns: 74/1; 126/1; 161/2; 221/1; 268/1; 347/2; 640/1; 820/3; 919/2

Query Match 28.0%; Score 109.5; DB 2; Length 949;  
 Best Local Similarity 39.0%; Pred. No. 0.0019;

Matches 23; Conservative 7; Mismatches 24; Indels 5; Gaps 3;

OY 12 CPSPNEIFSRCDRCORCPNVVPRKPLIC-1KICAPGCVCRGLGTLR--NKKKYCVPRSKC 66

Db 229 CGANEQYSACFSQSCOPSCOD-PSTPACAPACGCPGICLPGYIRDRSPRSACVPRGIC 286

RESULT 8  
 T34022

zonadhesin - pig

C:Species: Sus scrofa domestica (domestic pig)

C>Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000

C:Accession: T34022

R:Hardy, D.M.; Garbers, D.L.

J. Biol. Chem. 270, 26025-26028, 1995

A:Title: A sperm membrane protein that binds in a species-specific manner to the egg ext

A:Reference number: Z21464; MUID:96064658

A:Accession: T34022

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2476 <HAR>

A:Cross-references: EMBL:U40024; NID:g1066465; PID:g1066466; PIDN:AAC48486.1

A:Experimental source: strain Meishan; testis

C:Genetics:

A:Gene: zan

C:Function:

A:Description: may be involved in sperm adhesion to the zona pellucida

Query Match 27.5%; Score 107.5; DB 2; Length 2476;  
 Best Local Similarity 33.3%; Pred. No. 0.006;

Matches 22; Conservative 9; Mismatches 18; Indels 17; Gaps 3;

OY 11 KCPSPNEIFSRCDRC-----QRCFNPVPRKPLICIKICAPGCVCRGLGTLRNNKKKVCV 61

Db 1455 KCPSSGSSYTCANPCPATCTLNNPSPCTLP-----CAGCPCCKGHLISGTS-CV 1506

OY 62 PRSKC 67

Db 1507 PLSCQG 1512

RESULT 9  
 T15609

hypothetical protein C25E10.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T15609

R:Bradshaw, H.

submitted to the EMBL Data Library, February 1996

A:Description: The sequence of C. elegans cosmid C25E10.

A:Reference number: Z18376

A:Accession: T15609

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-137 <BRA>

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:C25E10.8

A:Introns: 1/3; 19/1; 42/3

Query Match 25.7%; Score 100.5; DB 2; Length 137;  
 Best Local Similarity 36.8%; Pred. No. 0.0034;

Matches 21; Conservative 7; Mismatches 26; Indels 3; Gaps 3;

OY 11 KCPSPNEIFSRCDRCORCPNVVPRKPLICIKICAPG-CVCRGLGTLRNNKKKVCVPRSKC 66

Db 81 KCPNTEFFRCGTACPEPCERKGRPR-CTRQCIYVWCGSSGFVRNGYR-CTELKEC 135

RESULT 10  
 T16574

hypothetical protein K05F1.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T16574

R:Wohlmann, P.

submitted to the EMBL Data Library, June 1995

A:Description: The sequence of C. elegans cosmid K05F1.

A:Reference number: Z18537

A:Accession: T16574

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-140 <WOB>

A:Cross-references: EMBL:U29377; NID:9868173; PID:9868180; PIDN:AAA68717.1; CESP:K05F

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:K05F1.10

A:Introns: 31/3; 77/2

Query Match 25.6%; Score 100; DB 2; Length 140;  
 Best Local Similarity 32.8%; Pred. No. 0.0039;

Matches 21; Conservative 10; Mismatches 25; Indels 8; Gaps 2;

OY 11 KCPSPNEIFSRCDRCORCPNVVPRKPLIC-----KICAPGCVCRGLGTLRNNKKKVCV 62

Db 52 ECKHEHHLICGPERHCDRTCEMLFSPHCLNHLHAKYCPRCVNDGVSRSEKICIR 111

OY 63 RSKC 66

Db 112 PSHC 115

RESULT 11  
 VMHU

von Willebrand factor precursor - human

C:Species: Homo sapiens (man)

C>Date: 04-Dec-1986 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999

C:Accession: A34480; S02377; A37139; S23676; A25298; A25469; A25366; S23645;

R:Mancuso, D.J.; Tulley, E.A.; Westfield, L.A.; Morrall, N.K.; Shelton-Inlles, B.B.; S

J. Biol. Chem. 264, 19514-19527, 1989

A:Title: Structure of the gene for human von Willebrand factor.

A:Reference number: A34480; MUID:90062044

A:Accession: A34480

A:Molecule type: DNA

A:Residues: 1-2813 <MAN>

A:Cross-references: EMBL:M25864

R:Bonthron, D.; Orkin, S.H.

Eur. J. Biochem. 171, 51-57, 1988

A:Title: The human von Willebrand factor gene. Structure of the 5' region.

A:Reference number: S02377; MUID:88111704

A:Accession: S02377

A:Molecule type: DNA

A:Residues: 1-177 <BO2>

A:Cross-references: EMBL:X06828

R:Mancuso, D.J.; Tulley, E.A.; Westfield, L.A.; Lester-Mancuso, T.L.; Le Beau, M.M.; S

A:Title: Human von Willebrand factor gene and pseudogene: structural analysis and differ  
A:Reference number: A37139; MUID:91105089  
A:Accession: A37139  
A:Molecule type: DNA  
A:Residues: 990-1947 <MAD>  
A:Cross-references: GB:M60675; NID:9340357; PIDN:AAA61295.1; PID:9553810  
A:Note: the authors translated the codon CGC for residue 156 as Gln  
R:Collins, C.J.; Underdahl, J.P.; Levene, R.B.; Revert, C.P.; Morin, M.J.; Dombalagian,  
Proc. Natl. Acad. Sci. U.S.A. 84, 4393-4397, 1987  
A:Title: Molecular cloning of the human gene for von Willebrand factor and identificatio  
A:Reference number: S23676; MUID:87260814  
A:Accession: S23676  
A:Molecule type: DNA  
A:Residues: 2731-2813 <COL>  
A:Cross-references: EMBL:M16945  
R:Bonthron, D.; Orr, E.C.; Mitscock, L.M.; Ginsburg, D.; Handin, R.I.; Orkin, S.H.  
Nucleic Acids Res. 14, 7125-7127, 1986  
A:Title: Nucleotide sequence of pre-pro-von Willebrand factor cDNA.  
A:Reference number: A25298; MUID:87016349  
A:Accession: A25298  
A:Molecule type: mRNA  
A:Residues: 1-470, 'V', 472-2813 <BON>  
A:Cross-references: EMBL:X04385  
R:Verweij, C.L.; Diergaarde, P.J.; Hart, M.; Pannekoek, H.  
EMBO J. 5, 1839-1847, 1986  
A:Title: Full-length von Willebrand factor (vWF) cDNA encodes a highly repetitive protei  
A:Reference number: A91044; MUID:87004550  
A:Accession: A25469  
A:Molecule type: mRNA  
A:Residues: 1-470, 'V', 472-483, 'R', 485-1022, 'K', 1024-1025, 'E', 1027-1400 <VER>  
A:Cross-references: EMBL:X0416  
A:Note: this sequence has been revised in reference A91056  
R:Verweij, C.L.; Diergaarde, P.J.; Hart, M.; Pannekoek, H.  
EMBO J. 5, 3074, 1986  
A:Reference number: A91056  
A:Accession: A25366  
A:Molecule type: mRNA  
A:Residues: 1021-1030 <VE2>  
A:Note: this is a revision to the sequence from reference A91044  
R:Shelton-Inloes, B.B.; Broze Jr., G.J.; Miletich, J.P.; Sadler, J.E.  
Biochem. Biophys. Res. Commun. 144, 657-665, 1987  
A:Title: Evolution of human von Willebrand factor: cDNA sequence polymorphisms, repeated  
A:Reference number: S23618; MUID:87213253  
A:Accession: S23618  
A:Molecule type: mRNA  
A:Residues: 1-120 <SH2>  
A:Cross-references: EMBL:M17588; NID:9799330; PIDN:AAA65940.1; PID:9340316  
A:Accession: S23645  
A:Molecule type: protein  
A:Residues: 23-56 <SH3>  
R:Sadler, J.E.; Shelton-Inloes, B.B.; Sorace, J.M.; Harlan, J.M.; Titani, K.; Davie, E.W.  
Proc. Natl. Acad. Sci. U.S.A. 82, 6394-6398, 1985  
A:Title: Cloning and characterization of two cDNAs coding for human von Willebrand facto  
A:Reference number: A94060; MUID:86016708  
A:Accession: A94060  
A:Molecule type: mRNA  
A:Residues: 'WA', 733, 'C', 744-769, 'H', 771-788, 'A', 790-803, 'S', 805-873, 1289-1471, 'D', 1473-  
A:Note: the authors translated the codon TGC for residue 168 as Cys  
R:Shelton-Inloes, B.B.; Titani, K.; Sadler, J.E.  
Biochemistry 25, 3164-3171, 1986  
A:Title: cDNA sequences for human von Willebrand factor reveal five types of repeated do  
A:Reference number: A90504; MUID:86269894  
A:Accession: A90504  
A:Molecule type: mRNA  
A:Residues: 781-788, 'A', 790-1424 <SHE>  
A:Note: 852-Gln, 857-Asp, and 1381-Thr were also found  
R:Ginsburg, D.; Handin, R.I.; Bonthron, D.T.; Donlon, T.A.; Bruns, G.A.P.; Latt, S.A.; C  
Science 228, 1401-1406, 1985  
A:Title: Human von Willebrand factor (vWF): isolation of complementary DNA (cDNA) clones  
A:Reference number: A44178; MUID:85244588  
A:Accession: A44178  
A:Molecule type: mRNA  
A:Residues: 3621-2813 <GIN>

A:Cross-references: EMBL:R03028; NID:9340308; PIDN:AAA61293.1; PID:9340309  
R:Verweij, C.L.; de Vries, C.J.M.; Distel, B.; van Zonneveld, A.J.; van Kessel, A.G.;  
Nucleic Acids Res. 13, 4699-4717, 1985  
A:Title: Construction of cDNA coding for human von Willebrand factor using antibody p  
A:Reference number: S07363; MUID:85269603  
A:Accession: S07363  
A:Molecule type: mRNA  
A:Residues: 2731-2813 <VE3>  
A:Cross-references: EMBL:X02672; NID:937939; PIDN:CAA2503.1; PID:937940  
R:Lynch, D.C.; Zimmerman, T.S.; Collins, C.J.; Brown, M.; Morin, M.J.; Ling, E.H.; Li  
Cell 41, 49-56, 1985  
A:Title: Molecular cloning of cDNA for human von Willebrand factor: authentication by  
A:Reference number: S23678; MUID:85201687  
A:Accession: S23678  
A:Molecule type: mRNA  
A:Residues: 2731-2813 <LYN>  
A:Cross-references: EMBL:R03028  
R:Titani, K.; Kumar, S.; Takio, K.; Ericsson, L.H.; Wade, R.D.; Ashida, K.; Walsh, K.  
Biochemistry 25, 3171-3184, 1986  
A:Title: Amino acid sequences of human von Willebrand factor.  
A:Reference number: A90505; MUID:86269895  
A:Accession: A90505  
A:Molecule type: protein  
A:Residues: 764-788, 'A', 790-1471, 'D', 1473-2813 <TIT>  
A:Note: 789-Thr was also found  
R:Chopek, M.W.; Girma, J.P.; Fujikawa, K.; Davie, E.W.; Titani, K.  
Biochemistry 25, 3146-3155, 1986  
A:Title: Human von Willebrand factor: a multivalent protein composed of identical sub  
A:Reference number: A23464; MUID:86269892  
A:Accession: A23464  
A:Molecule type: protein  
A:Residues: 764-773; 2803-2813 <CHO>  
R:Dent, J.A.; Berkowitz, S.D.; Ware, J.; Kasper, C.K.; Ruggeri, Z.M.  
Proc. Natl. Acad. Sci. U.S.A. 87, 6306-6310, 1990  
A:Title: Identification of a cleavage site directing the immunochemical detection of  
A:Reference number: A36013; MUID:90349604  
A:Accession: A36013  
A:Molecule type: protein  
A:Residues: 1606-1617 <DEN>  
R:Fay, P.J.; Kawai, Y.; Wagner, D.D.; Ginsburg, D.; Bonthron, D.; Ohlsson-Wilhelm, B.  
Science 232, 995-998, 1986  
A:Title: Propolypeptide of von Willebrand factor circulates in blood and is identical  
A:Reference number: A60913; MUID:86208144  
A:Accession: A60913  
A:Molecule type: protein  
A:Residues: 576-590 <FAY>  
A:Gene: GDB:VWF  
A:Cross-references: GDB:119125; OMIM:193400  
A:Map position: 12p13.3-12p13.2  
A:Introns: 19/1; 74/1; 108/2; 178/1; 219/3; 292/1; 333/1; 370/2; 386/1; 431/3; 478/1;  
5/1; 1724/1; 1711/1; 1819/1; 1874/1; 1888/3; 1948/1; 2021/3; 2086/1; 2200/1; 2266/3;  
C:Superfamily: von Willebrand factor: von Willebrand factor type A repeat homology; v  
C:Keywords: blood coagulation; cell binding; connective tissue; disulfide bond; dup  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-763/Domain: von Willebrand antigen II #status predicted <WAI>  
F:334-386/Domain: type D repeat 1 <DD1>  
F:387-745/Domain: type D repeat 2 <DD2>  
F:766-700/Region: cell attachment (R-G-D) motif  
F:764-2813/Domain: von Willebrand factor #status predicted <MA2>  
F:784-865/Domain: D <DDD>  
F:788-833, 2216-2261/Region: duplication  
F:826-853, 2400-2515, 2544-2662/Region: duplication  
F:842-1130, 1934-2203/Region: duplication  
F:866-1241/Domain: type D repeat 3 <DD3>  
F:1275-1443/Domain: von Willebrand factor type A repeat homology <WAI1>  
F:1466-1654/Domain: von Willebrand factor type A repeat homology <WAI2>  
F:1689-1854/Domain: von Willebrand factor type A repeat homology <WAI3>  
F:1947-2295/Domain: type D repeat 4 <DD4>  
F:2296-2330/Domain: type B repeat 1 <VB1>  
F:2340-2365/Domain: type B repeat 2 <VB2>  
F:2375-2399/Domain: type B repeat 3 <VB3>  
F:2430-2497/Domain: von Willebrand factor type C repeat homology <WVIC1>

F:2507-2509/Region: cell attachment (R-G-D) motif  
 F:2581-2647/Domain: von Willebrand factor type C repeat homology <VMC2>  
 F:857,1231,1515,1574,2223,2290,2357,2400,2546,2585,2790/Binding site: carbohydrate (Asn)  
 F:1147/Binding site: carbohydrate (Asn) (covalent) #status atypical  
 F:1248,1255,1256,1468,1477,1487,1679,2298/Binding site: carbohydrate (Thr) (covalent) #s  
 F:1263,1486/Binding site: carbohydrate (Ser) (covalent) #status experimental

Query Match 25.1%; Score 96; DB 1; Length 2813;  
 Best Local Similarity 26.0%; Pred. No. 0.056;  
 Matches 19; Conservative 13; Mismatches 27; Indels 14; Gaps 2;

OY 8 GRC-----KCPSEIFSRDGRQRCPCPNVVPKPLCTIKICAPGCVCRLGVL 53  
 DB 635 GRGVRVAMREPCRELCNCPKGVYLCQCTPCNLTRSLSTPEECNEALSCFCPPGLR 694

OY 54 RNKKVCVPRSKC 66  
 DB 695 MDERGDCVPRKAC 707

RESULT 12  
 T30243

alpha tectorin - chicken  
 C:Species: Gallus gallus (chicken)

C>Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000

C:Accession: T30243

R:Containo, P.; Goodyear, R.G.; Legan, P.K.; Richardson, G.P.

Hear. Res. 130, 62-74, 1999

A:Title: Chick alpha tectorin : molecular cloning and expression during embryogenesis.

A:Reference number: 220783; MUID:99251817

A:Accession: T30243

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2120 <COO>

A:Cross-references: EMBL:AJ012287; NID:e1361091; PID:e1361092; PIDN:CAA09979.1

A:Note: non-collagenous protein only expressed in the inner ear

Query Match 24.9%; Score 97.5; DB 2; Length 2120;  
 Best Local Similarity 35.7%; Pred. No. 0.051;  
 Matches 20; Conservative 6; Mismatches 29; Indels 1; Gaps 1;

OY 12 CPSEIFSRDGRQRCPCPNVVPKPLCTIKICAPGCVCRLGVLRNKKVCVPRSKC 67  
 DB 1345 CPNHSYEVSCVLCQPCRAIRLSDCGHYCVGCGCCDGVLYLNKSKLIPON-CG 1399

RESULT 13  
 T17405

scavenger receptor cysteine-rich protein precursor - sea urchin (Strongylocentrotus purpuratus)  
 C:Species: Strongylocentrotus purpuratus (purple urchin)

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-May-2000

C:Accession: T17405

R:Rancier, Z.; Rast, J.P.; Davidson, E.H.

Immunogenetics 49, 773-786, 1999

A:Title: Origins of immunity: transcription factors and homologs of effector genes of th

A:Reference number: 218253; MUID:99328904

A:Accession: T17405

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1036 <PAP>

A:Cross-references: EMBL:AF076513; NID:g4164530; PID:g4164531; PIDN:AAD05493.1

Query Match 24.7%; Score 96.5; DB 2; Length 1036;  
 Best Local Similarity 30.6%; Pred. No. 0.038;  
 Matches 19; Conservative 10; Mismatches 30; Indels 3; Gaps 2;

OY 8 GRCPCPSNEIFSRDGRQRCPCPNVVPKPLCTIKICAPGCVCRLGVLRNKK--KVCVPRSK 64  
 DB 555 GDHHCPEGMTEFNEGSGCGGSCDNLVPRDLCPLFCFVGCFEGIVKQDGGDGRCPVD 614

OY 65 KC 66  
 DB 615 QC 616

RESULT 14  
 T30257

IgG Fc binding protein - mouse (fragment)

C:Species: Mus musculus (house mouse)

C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999

C:Accession: T30257

R:Tomasetto, C.; Masson, R.; Wendling, C.; Lefebvre, O.; Chenard, M.P.; Ribieras, S.;

submitted to the EMBL Data Library, September 1998

A:Description: Identification of interactions between trefoil peptides and members of

A:Reference number: 220795

A:Accession: T30257

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1023 <TOM>

A:Cross-references: EMBL:AJ011416; NID:e1325179; PID:e1325180; PIDN:CAA09618.1

Query Match 24.2%; Score 94.5; DB 2; Length 1023;  
 Best Local Similarity 35.1%; Pred. No. 0.059;  
 Matches 20; Conservative 7; Mismatches 29; Indels 1; Gaps 1;

OY 11 KCPSEIFSRDGRQRCPCPNVVPKPLCTIKICAPGCVCRLGVLRNKKVCVPRSKC 67  
 DB 229 ECPHSYEVSCVLCQPCRAIRLSDCGHYCVGCGCCDGVLYLNKSKLIPON-CG 264

RESULT 15  
 T25513

hypothetical protein C04E6.12 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000

C:Accession: T25513

R:Bentley, D.; Gattung, S.

submitted to the EMBL Data Library, April 1997

A:Description: The sequence of C. elegans cosmid C04E6.

A:Reference number: 220043

A:Accession: T25513

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-192 <BEN>

A:Cross-references: EMBL:U07012; PIDN:AAB52332.1; GSPDB:GN00023; CESP:C04E6.12

A:Experimental source: strain Bristol N2; clone C04E6

C:Genetics:

A:Gene: CESP:C04E6.12

A:Map position: 5

C:Superfamily: Caenorhabditis elegans hypothetical protein C04E6.12

Query Match 23.8%; Score 93; DB 2; Length 192;  
 Best Local Similarity 28.6%; Pred. No. 0.024;  
 Matches 22; Conservative 8; Mismatches 21; Indels 26; Gaps 3;

OY 12 CPSEIFSRDGRQRCPCPNVVPKPLCTIKI-----CAPGCYCR 49  
 DB 88 GSKTEIYHCLD--CEPTCHNLIIPK--CRKVRTEATVETIRDTRVRSVBOCKMKGCVC 143  
 OY 50 LGLRNKKKKVCVPRSKC 66  
 DB 144 TGLARNABGKCVTLREC 160

Search completed: June 21, 2001, 10:51:02  
 Job time: 210 sec

Thu Jun 21 13:32:07 2001

us-09-506-978-1.rpr

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2001, 10:50:47 ; Search time 21.34 Seconds  
(without alignments)  
415.391 million cell updates/sec

Title: US-09-506-978-1  
Perfect score: 391  
Sequence: 1 GGFGGIGRGKCPSENEIFSR.....CRIGYLNKKKVCVRSKCG 67

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 133305027 residues  
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP unclassified:\*  
13: SP vertebrate:\*  
14: SP virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	133	34.0	245	5	Q19964
2	118	30.2	561	5	Q9U1T6
3	118	30.2	626	5	Q9U1T5
4	113	28.9	2155	11	Q08523
5	113	28.9	5376	11	Q88799
6	111.5	28.5	166	5	Q16488
7	109.5	28.0	949	5	P90956
8	109	27.9	2379	4	Q9Y493
9	108	27.6	3843	5	Q9U5D0
10	108	27.6	3843	5	Q9YU94
11	107	27.4	2155	4	Q75443
12	106.5	27.2	2843	4	Q9Y6R7
13	104.5	26.7	453	5	Q9U1U6
14	100.5	25.7	137	5	Q18157
15	100.5	25.7	211	5	Q9U1U0
16	100.5	25.7	735	4	Q95784
17	100	25.6	91	5	Q16938
18	100	25.6	140	5	Q21248
19	99	25.3	92	5	Q9GPC4

20	97.5	24.9	2120	13	Q9YH85	Q9YH85 gallus gall
21	96.5	24.7	505	5	Q25431	Q25431 lytechinus
22	96.5	24.7	1036	5	Q97378	Q97378 strongyloce
23	94.5	24.2	98	5	Q94162	Q94162 caenorhabdi
24	93	23.8	192	5	Q01471	Q01471 caenorhabdi
25	93	23.8	1637	6	Q9XSV8	Q9XSV8 bos taurus
26	92.5	23.7	108	5	Q3VUW1	Q3VUW1 bos taurus
27	92.5	23.7	195	5	Q18805	Q18805 caenorhabdi
28	92	23.5	249	5	Q45764	Q45764 caenorhabdi
29	92	23.5	648	5	Q9NKD7	Q9NKD7 drosophila
30	92	23.5	701	5	Q9VUD4	Q9VUD4 drosophila
31	91	23.3	92	5	Q9GPC8	Q9GPC8 drosophila
32	91	23.3	1104	4	Q60460	Q60460 homo sapien
33	91	23.3	1373	4	Q75372	Q75372 homo sapien
34	90.5	23.1	135	5	Q18158	Q18158 caenorhabdi
35	90	23.0	457	5	Q9NC91	Q9NC91 strongyloce
36	90	23.0	2327	13	Q9IBG7	Q9IBG7 xenopus lae
37	90	23.0	4123	4	Q75851	Q75851 homo sapien
38	89.5	22.9	1642	5	Q62055	Q62055 caenorhabdi
39	88.5	22.6	62	5	Q77419	Q77419 ascaris suu
40	88.5	22.6	145	5	Q18156	Q18156 caenorhabdi
41	88.5	22.6	316	14	Q57092	Q57092 ectromelia
42	88.5	22.6	320	14	Q57091	Q57091 ectromelia
43	88.5	22.6	320	14	Q57300	Q57300 ectromelia
44	88.5	22.6	513	11	Q08647	Q08647 mus musculu
45	88	22.5	92	5	Q9GPC7	Q9GPC7 drosophila

## ALIGNMENTS

RESULT 1  
ID Q19964 PRELIMINARY; PRT; 245 AA.  
AC Q19964;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)  
DE F32D8.3 PROTEIN.  
GN F32D8.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Wilkinson J.;  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,  
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin K., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
DR EMBL; Z74031; CAA98455.1; -;  
DR HSSP; P56682; ICCV.  
DR InterPro; IPR000561; -;  
DR InterPro; IPR002919; -;  
DR Pfam; PF01826; TIL; 1.  
DR PROSITE; PS0186; EGF\_2; UNKNOWN\_1.  
SQ SEQUENCE 245 AA: 27785 MW: 781AEAV7F9FE784C CRC64;

Query Match 34.0%; Score 133; DB 5; Length 245;  
 Best Local Similarity 43.1%; Pred. No. 1.2e-08;  
 Matches 28; Conservative 7; Mismatches 20; Indels 10; Gaps 4;

QY 6 LGGRCGKPSNIFSRCDGRC-QRFCEPNV-VPKPLCIKICAPGCVCRGLYLRNKKKVCVPRSKC 66  
 DB 57 INGENCPCMFQSFHC--ACESTCNN--PDPTCSK-CEGCTCRNGFVNSLKLCVLPPEE 111  
 QY 62 -PRSK 65  
 DB 112 CPRTK 116

RESULT 2  
 ID 0901T6 PRELIMINARY; PRT; 561 AA.  
 AC 0901T6;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE Y69H2.3A PROTEIN.  
 GN Y69H2.3A.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA McMurray A.A.;  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology.";  
 RT Science 282:2012-2018(1998).  
 RL EMBL: 298877; CAB54472.1; -;  
 DR HSSP: P56682; ICCV.  
 DR InterPro: IPR000561; -;  
 DR InterPro: IPR002919; -;  
 DR Pfam: PF01826; TIL; 3.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN\_3.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN\_3.  
 SQ SEQUENCE 561 AA; 60751 MW; 46930233355972B2 CRC64;

Query Match 30.2%; Score 118; DB 5; Length 561;  
 Best Local Similarity 36.8%; Pred. No. 1.6e-06;  
 Matches 21; Conservative 11; Mismatches 23; Indels 2; Gaps 2;

QY 12 CPSNIFSRCDGRC-QRFCEPNV-VPKPLCIKICAPGCVCRGLYLRNKKKVCVPRSKC 66  
 DB 123 CPVNEVSNCHNPTCEKCKPQKNAPQVNCIMACQVCGSCMDGFVRNNGVCVKEAEC 179

RESULT 3  
 ID 0901T5 PRELIMINARY; PRT; 626 AA.  
 AC 0901T5;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE Y69H2.3B PROTEIN.  
 GN Y69H2.3B.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA McMurray A.A.;  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology.";  
 RT Science 282:2012-2018(1998).  
 RL EMBL: 298877; CAB54473.1; -;  
 DR HSSP: P56682; ICCV.  
 DR InterPro: IPR000561; -;  
 DR InterPro: IPR002919; -;  
 DR Pfam: PF01826; TIL; 4.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN\_3.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN\_3.  
 SQ SEQUENCE 626 AA; 67881 MW; A17F7B445800E118 CRC64;

Query Match 30.2%; Score 118; DB 5; Length 626;  
 Best Local Similarity 36.8%; Pred. No. 1.8e-06;  
 Matches 21; Conservative 11; Mismatches 23; Indels 2; Gaps 2;

QY 12 CPSNIFSRCDGRC-QRFCEPNV-VPKPLCIKICAPGCVCRGLYLRNKKKVCVPRSKC 66  
 DB 123 CPVNEVSNCHNPTCEKCKPQKNAPQVNCIMACQVCGSCMDGFVRNNGVCVKEAEC 179

RESULT 4  
 ID 008523 PRELIMINARY; PRT; 2155 AA.  
 AC 008523;  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE TECTORIN ALPHA (ALPHA TECTORIN).  
 GN TECTA.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CD1; TISSUE=WHOLE COCHLEAE;  
 RX MEDLINE=97236843; PubMed=9079715;  
 RA Legan P.K., Rau A., Keene J.N., Richardson G.P.;  
 RT "The mouse tectorins. Modular matrix proteins of the inner ear  
 RT homologous to components of the sperm-egg adhesion system.";  
 RT J. Biol. Chem. 272:8791-8801(1997).  
 RL EMBL: X99805; CAA68138.1; -;  
 DR HSSP: P56682; ICCV.  
 DR HSSP: MGI:109575; Tecta.  
 DR InterPro: IPR000421; -;  
 DR InterPro: IPR000561; -;  
 DR InterPro: IPR001507; -;  
 DR InterPro: IPR001846; -;  
 DR InterPro: IPR002086; -;  
 DR InterPro: IPR002919; -;  
 DR Pfam: PF00094; vwd; 4.  
 DR Pfam: PF00100; zona\_pellucida; 1.  
 DR Pfam: PF01826; TIL; 3.  
 DR PROSITE: PS00070; ALDEHYDE\_DEHYDR\_CYS; UNKNOWN\_1.  
 DR PROSITE: PS01186; EGF\_2; 3.  
 DR PROSITE: PS01286; FA58C\_2; UNKNOWN\_1.  
 DR PROSITE: PS00682; ZP\_DOMAIN; 1.  
 DR SMART: SM00001; EGF\_Like; 1.  
 KW EGF-like domain; Glycoprotein.  
 SQ SEQUENCE 2155 AA; 239534 MW; 86C5C09AA392B1C5 CRC64;

Query Match 28.9%; Score 113; DB 11; Length 2155;  
 Best Local Similarity 38.6%; Pred. No. 2.1e-05;  
 Matches 22; Conservative 7; Mismatches 26; Indels 2; Gaps 2;

QY 11 KPSNIFSRCDGRC-QRFCEPNV-VPKPLCIKICAPGCVCRGLYLRNKKKVCVPRSKC 67  
 DB 983 ECPENHFECC-MTCTETETLALGPICVDSCEGCGDEGY-RLGSGCVTRSECG 1037

```

RESULT 5
088799 PRELIMINARY; PRT; 5376 AA.
AC 088799;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE ZONADHESIN.
GN ZAN
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98123114; PubMed=9452463;
RA Gao Z., Garbers D.L.;
RT "Species diversity in the structure of zonadhesin, a sperm-specific
RT membrane protein containing multiple cell adhesion molecule-like
RT domains."
RL J. Biol. Chem. 273:3415-3421(1998).
DR EMBL; U97068; AAC26880.1; -.
DR InterPro: IPR000561; -.
DR InterPro: IPR000998; -.
DR InterPro: IPR001846; -.
DR InterPro: IPR002919; -.
DR InterPro: IPR003328; -.
DR Pfam: PF00094; vwd; 4.
DR Pfam: PF00629; MAM; 3.
DR Pfam: PF01826; TIL; 25.
DR Pfam: PF02345; Tila; 25.
DR PROSITE: PS00022; EGF_2; 18.
DR PROSITE: PS01186; EGF_2; 18.
DR PROSITE: PS50050; MAM_2; 3.
DR SMART; SM00001; EGF_Like; 1.
DR EGF-like domain; Glycoprotein.
SQ SEQUENCE 5376 AA; 579908 MW; 0E44DB77DF2A2620 CRC64;

Query Match 28.9%; Score 113; DB 11; Length 5376;
Best Local Similarity 37.5%; Pred. No. 4.6e-05;
Matches 24; Conservative 8; Mismatches 22; Indels 10; Gaps 3;

QY 11 KCPSNEIFSRGRCQRCFCPNVVPKPLC-----IKICAPGVCYCRGILTKKKKVCVPR 63
DB 4743 KCRANSIYTHCLPTCLPSCSN--PDGRCEGTSKAPSTCREGCVCPGYILN-KDTCVHK 4799
QY 64 SKCG 67
DB 4800 NQCG 4803

RESULT 6
016488 PRELIMINARY; PRT; 166 AA.
AC 016488;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE B0238.12 PROTEIN.
GN B0238.12
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

```

```

RA Craxton M., Dear S., Du Z., Durbin R., Favella A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopay A., Saunders D., Shownkeen R.,
RA Snelson N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Blanchard M., Bradshaw H.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF016450; AAB65990.1; -.
DR HSSP; P56682; ICGV.
DR InterPro: IPR000561; -.
DR InterPro: IPR002919; -.
DR Pfam; PF01826; TIL; 2.
DR PROSITE; PS01186; EGF_2; UNKNOWN_2.
SQ SEQUENCE 166 AA; 17971 MW; DA92F2473442ABD CRC64;

Query Match 28.5%; Score 111.5; DB 5; Length 166;
Best Local Similarity 36.8%; Pred. No. 3.6e-06;
Matches 25; Conservative 11; Mismatches 25; Indels 7; Gaps 4;

QY 4 GGIGGRGKCP---SNEIFSRGRCQRCFCPNVVPKPLKIKIAPG-CYCRGILNKK-KK 58
DB 27 GOYGGGQRLPCRCGRNDEYVTCGTACEPSCN--PNPMCTKOCINNVCCOCSGVYNEITR 84
QY 59 VCVPRSK 66
DB 85 QCVROAQC 92

RESULT 7
P90956 PRELIMINARY; PRT; 949 AA.
AC P90956;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE T01D3.3 PROTEIN.
GN T01D3.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC Steward C.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) = O(2) + H(2)O(2)
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
DR EMBL; Z81110; CAB03260.1; -.
DR EMBL; Z81110; CAB03259.1; -.
DR HSSP; P56682; ICGV.
DR InterPro: IPR000716; -.
DR InterPro: IPR001424; -.
DR InterPro: IPR002221; -.
DR InterPro: IPR002919; -.
DR Pfam; PF00080; sodcu; 1.

```

DR Pfam: PF00086; thyroglobulin\_1, 1.  
 DR Pfam: PF01826; TIL, 6.  
 DR PROSITE: PS00317; 4-DISULFIDE\_CORE; 1.  
 DR SMART: SM00211; TY; 1.  
 KW Alternative splicing; Copper; Oxidoreductase; Zinc.  
 SQ SEQUENCE 949 AA; 103486 MW; C850B2886C98EF5 CRC64;

Query Match 28.0%; Score 109.5; DB 5; Length 949;  
 Best Local Similarity 39.0%; Pred. No. 2.8e-05;  
 Matches 23; Conservative 7; Mismatches 24; Indels 5; Gaps 3;

OY 12 CPSENEIFSRCDGRCPNVVPRKPLCITKICAGCVCRLGLYLNKKNVCYPRSKC 66  
 DB 229 CGANQYSAFCSSCPSCOD-PSPPACAPACQPCICLPYIRKDSPPRSACVPRGLC 286

## RESULT 8

OY9493 PRELIMINARY; PRT: 2379 AA.

AC OY9493:  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE ZONADHESIN (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Tsui L.-C.; Scherer S.; Weber J.; Schattevoy R.,  
 RA Gloeckner G.; Renenthal A.; Scherer S.; Weber J.; Schattevoy R.,  
 RT "Large scale analysis of two regions in human chromosome 7q22:  
 RT annotation of 650 kb of genomic sequence around the PCOLCE and CUTL1  
 RT loci reveals 17 genes."  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF053356; AAC78790.1; -  
 DR InterPro: IPR000561; -  
 DR InterPro: IPR000988; -  
 DR InterPro: IPR001846; -  
 DR InterPro: IPR002919; -  
 DR InterPro: IPR002965; -  
 DR InterPro: IPR003328; -  
 DR Pfam: PF00094; vwd; 4.  
 DR Pfam: PF00629; MAM; 3.  
 DR Pfam: PF01826; TIL; 4.  
 DR Pfam: PF02345; TIL; 4.  
 DR PRINTS: PR01217; PRICHEXTENS.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN\_3.  
 DR PROSITE: PS00740; MAM\_1; 1.  
 DR PROSITE: PS50060; MAM\_2; 6.  
 DR SMART: SM00216; VWD; 1.  
 FT NON\_TER 2379  
 FT 2379  
 SQ SEQUENCE 2379 AA; 257820 MW; 690D092316CEC94D CRC64;

Query Match 27.9%; Score 109; DB 4; Length 2379;  
 Best Local Similarity 35.3%; Pred. No. 7e-05;  
 Matches 24; Conservative 11; Mismatches 13; Indels 20; Gaps 4;

OY 11 KCPSNEIFSRCDGRCPNVVPRKPLCITKICAGCVCRLGLYLNKKNVCYPRSKC 59  
 DB 2173 ECRAYSITMCLPSCPSCHDLGRCE--GAKVP-----SACAEGCICQPGVISEDK- 2223

OY 60 CVPRSKCG 67  
 DB 2224 CVPRSKCG 2231

RESULT 9  
 OY95D0 PRELIMINARY; PRT: 3843 AA.

AC OY95D0:  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE HEMOLECTIN.  
 GN HML OR CG7002.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Goto A.; Kumagai C.; Kumagai T.; Hirose J.; Narita H.; Beck K.,  
 RA Mori H.; Kadowaki T.; Kitagawa Y.;  
 RT "Identification and cloning of Hemolactin: A Drosophila protein with  
 RT homology to the silkorm hemocytin and mammalian von Willebrand  
 RT factor."  
 RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A  
 CC (LDLRA) DOMAIN.  
 DR EMBL: AB035891; BAA8518.1; -  
 DR HSSP: P56682; 1CV.  
 DR Flybase: FBgn0029167; Hml.  
 DR InterPro: IPR000359; -  
 DR InterPro: IPR000421; -  
 DR InterPro: IPR000436; -  
 DR InterPro: IPR000561; -  
 DR InterPro: IPR001007; -  
 DR InterPro: IPR001092; -  
 DR InterPro: IPR001846; -  
 DR InterPro: IPR002172; -  
 DR InterPro: IPR002919; -  
 DR Pfam: PF00094; vwd; 5.  
 DR Pfam: PF00754; F5\_F8\_Type\_C; 2.  
 DR Pfam: PF01826; TIL; 6.

DR PROSITE: PS01185; CTCK\_1; UNKNOWN\_1.  
 DR PROSITE: PS00022; CTCK\_2; 1.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_2.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.  
 DR PROSITE: PS01286; FAS8C\_2; UNKNOWN\_1.  
 DR PROSITE: PS50068; LDLRA\_2; 1.  
 DR PROSITE: PS01208; VWF; UNKNOWN\_1.  
 DR SMART: SM00032; CCP; 1.  
 DR GlycoProtein.  
 KW Glycoprotein.  
 SQ SEQUENCE 3843 AA; 426349 MW; 20095BC3BF21E49D CRC64;

Query Match 27.6%; Score 108; DB 5; Length 3843;  
 Best Local Similarity 33.9%; Pred. No. 0.00014;  
 Matches 19; Conservative 7; Mismatches 30; Indels 0; Gaps 0;

OY 11 KCPSNEIFSRCDGRCPNVVPRKPLCITKICAGCVCRLGLYLNKKNVCYPRSKC 66  
 DB 1130 KCPGLGVFDECDGDCALSCDDLPSKSGCRPCVGHGVEYNEDGECVPRKMC 1185

## RESULT 10

OY9V94 PRELIMINARY; PRT: 3843 AA.

AC OY9V94:  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE HML PROTEIN.  
 GN HML OR CG7002.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.



RC STRAIN-BERKELEY;  
 RX MEDLINE-20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe C.R., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bertman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burks K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Kelthum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,  
 RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitzkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A  
 CC (LDLRA) DOMAIN.  
 CC EMBL: AE003536; AAF49795.2; -  
 DR HSSP: P56682; ICSV.  
 DR Flybase: FBgn0029167; Hml.  
 DR InterPro: IPR000359; -  
 DR InterPro: IPR000421; -  
 DR InterPro: IPR000561; -  
 DR InterPro: IPR001007; -  
 DR InterPro: IPR001846; -  
 DR InterPro: IPR002172; -  
 DR InterPro: IPR002919; -  
 DR Pfam: PF00094; vwd; 10.  
 DR Pfam: PF00754; F5\_F8\_Type\_C; 4.  
 DR Pfam: PF01826; TIL; 14.  
 DR PROSITE: PS01185; CTCK\_1; UNKNOWN\_1.  
 DR PROSITE: PS01225; CTCK\_2; 1.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_2.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.  
 DR PROSITE: PS01286; FA58C\_2; UNKNOWN\_1.  
 DR PROSITE: PS00682; LDLRA\_2; 1.  
 DR PROSITE: PS01208; VMEC; UNKNOWN\_1.  
 DR SMART: SM00001; EGF\_like; 1.  
 KW Glycoprotein.  
 SQ SEQUENCE 3843 AA; 426241 MW; 78D770C0BB6319D CRC64;

Query Match 27.6%; Score 108; DB 5; Length 3843;  
 Best Local Similarity 33.9%; Pred. No. 0.0014;  
 Matches 19; Conservative 7; Mismatches 30; Indels 0; Gaps 0;  
 Oy 11 KCPNSNIFRCGRCORCPNVVPPPLCIKICAPGCVCRGLGRNKKKVCYVPRSCG 66  
 Db 1130 KCPNGVFECDGCGALSCDILPSKSGCKRECVCRCRPHGVNEDCDGVCYKMC 1185

RESULT 11  
 ID 075443 PRELIMINARY; PRT; 2155 AA.  
 AC 075443  
 DT 01-NOV-1998 (TREMBLrel. 08. Created)  
 DT 01-NOV-1998 (TREMBLrel. 08. Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16. Last annotation update)  
 DE ALPHA-TECTORIN.  
 GN TECTA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RX MEDLINE=98250172; PubMed=9590290;  
 RP SEQUENCE FROM N.A.  
 RA Verhoeven K., Van Laer L., Kirschhofer K., Legan P.K., Hughes D.C.,  
 RA Schattelman I., Verstrecken M., Van Hauwe P., Coucke P., Chen A.,  
 RA Smith R.J.H., Somers T., Officiers F.E., Van de Heyning P.,  
 RA Goverts P.J., Van Camp G.,  
 RT "Mutations in the human alpha-*tektorin* gene cause autosomal dominant  
 non-syndromic hearing impairment.";  
 RL Nat. Genet. 19:60-62(1998).  
 DR EMBL: AF055136; AAC26019.1; JOINED.  
 DR EMBL: AF055114; AAC26019.1; JOINED.  
 DR EMBL: AF055115; AAC26019.1; JOINED.  
 DR EMBL: AF055116; AAC26019.1; JOINED.  
 DR EMBL: AF055117; AAC26019.1; JOINED.  
 DR EMBL: AF055118; AAC26019.1; JOINED.  
 DR EMBL: AF055119; AAC26019.1; JOINED.  
 DR EMBL: AF055120; AAC26019.1; JOINED.  
 DR EMBL: AF055121; AAC26019.1; JOINED.  
 DR EMBL: AF055122; AAC26019.1; JOINED.  
 DR EMBL: AF055123; AAC26019.1; JOINED.  
 DR EMBL: AF055124; AAC26019.1; JOINED.  
 DR EMBL: AF055125; AAC26019.1; JOINED.  
 DR EMBL: AF055126; AAC26019.1; JOINED.  
 DR EMBL: AF055127; AAC26019.1; JOINED.  
 DR EMBL: AF055128; AAC26019.1; JOINED.  
 DR EMBL: AF055129; AAC26019.1; JOINED.  
 DR EMBL: AF055130; AAC26019.1; JOINED.  
 DR EMBL: AF055131; AAC26019.1; JOINED.  
 DR EMBL: AF055132; AAC26019.1; JOINED.  
 DR EMBL: AF055133; AAC26019.1; JOINED.  
 DR EMBL: AF055134; AAC26019.1; JOINED.  
 DR EMBL: AF055135; AAC26019.1; JOINED.  
 DR HSSP: P56682; ICSV.  
 DR InterPro: IPR000421; -  
 DR InterPro: IPR000561; -  
 DR InterPro: IPR001507; -  
 DR InterPro: IPR001846; -  
 DR InterPro: IPR002919; -  
 DR Pfam: PF00094; vwd; 4.  
 DR Pfam: PF00100; zona\_pellucida; 1.  
 DR Pfam: PF01826; TIL; 3.  
 DR PROSITE: PS01186; EGF\_2; 3.  
 DR PROSITE: PS01286; FA58C\_2; UNKNOWN\_1.  
 DR PROSITE: PS00682; ZP\_DOMAIN; 1.  
 DR SMART: SM00001; EGF\_like; 1.  
 KW EGF-like domain; Glycoprotein.  
 SQ SEQUENCE 2155 AA; 239425 MW; 54EAB812AE3690C9 CRC64;

Query Match 27.4%; Score 107; DB 4; Length 2155;  
 Best Local Similarity 36.8%; Pred. No. 0.00011;  
 Matches 21; Conservative 7; Mismatches 27; Indels 2; Gaps 2;  
 Oy 11 KCPNSNIFRCGRCORCPNVVPPPLCIKICAPGCVCRGLGRNKKKVCYVPRSCG 67  
 Db 983 ECPNSHFEECI-TCTETCELTGIPICVDSGSGCCDCGVALLSQ-CVTRSECG 1037

```

RESULT 12
Q9Y6R7 PRELIMINARY; PRT; 2843 AA.
AC Q9Y6R7;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE HUMAN FC GAMMA BP (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCreedy P.M., Richardson P., Sakalasis G.,
RA Burkhart-Schultz K., Gordon L., Dias J., Scott D., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Skowronski E.,
RA Dangnan L., Erlar A., Christensen M., Georgescu A., Avila J.,
RA Johnson G., Attix C., Andreise T., Amico-Keller G., Coefield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,
RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carrano A.V.;
RT "Sequence analysis of a 2.3 Mb region in 19q13.1 containing the RYR
RT gene."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC007842; AAD39266.1; -
DR InterPro: IPR000561; -
DR InterPro: IPR001846; -
DR InterPro: IPR002045; -
DR InterPro: IPR002919; -
DR InterPro: IPR003645; -
DR Pfam: PF00094; vwd; 6.
DR Pfam: PF01826; TIL; 6.
DR PRINTS: PR00858; MTCRUSTACEAN.
DR PROSITE: PS01186; EGF_2; 6.
DR SMART: SM00274; FOLN; 1.
DR EGF-like domain; Glycoprotein.
FT NON_TER 2843 2843
SQ SEQUENCE 2843 AA; 301805 MW; D2C35B53281E5269 CRC64;

Query Match 27.2%; Score 106.5; DB 4; Length 2843;
Best Local Similarity 36.8%; Pred. No. 0.00017;
Matches 21; Conservative 6; Mismatches 29; Indels 1; Gaps 1;

QY 11 KCPSENFSCDRC-QRCPCPNVVPKPLCIKICAPGCVCRIGLRNKKVCVPSKCG 67
DB 1531 ECPNHSYELCADTCSLGCASLAPQCQDGCACGCGCDSGFLYN-GOACVPIQCG 1586

RESULT 13
Q9U106 PRELIMINARY; PRT; 453 AA.
AC Q9U106;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE Y69H2.4 PROTEIN.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EGF-like domain; Glycoprotein.
RX MEDLINE=99069613; Pubmed=9851916;
RA none;

```

```

RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018.(1998).
DR EMBL: Z98877; CAB63401.1; -
DR InterPro: IPR000561; -
DR InterPro: IPR002919; -
DR Pfam: PF01826; TIL; 4.
DR PROSITE: PS01186; EGF_2; UNKNOWN-3.
SQ SEQUENCE 453 AA; 47861 MW; FD4CC2A42CDB8DC4 CRC64;

Query Match 26.7%; Score 104.5; DB 5; Length 453;
Best Local Similarity 36.5%; Pred. No. 6e-05;
Matches 23; Conservative 6; Mismatches 25; Indels 9; Gaps 3;

QY 12 CPSNEIFSRCDRC-QRCPCPNVVPKPLCIKICAPGCVCRIGLRN-----KKKVCVPS 64
DB 215 CARNOTMSDCLNTCSDEKPCGKSMKCTKHGGGACACASGYLSSDGEYKPKDPP-- 272

QY 65 KCG 67
DB 273 ECG 275

RESULT 14
Q18157 PRELIMINARY; PRT; 137 AA.
AC Q18157;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE SIMILARITY TO EGF-LIKE DOMAIN CYSTEINE PATTERN SIGNATURE.
GN C25E10.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; Pubmed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Lalster N., Latteille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Patsion J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
RA Smaison N., Smith A., Sonnenhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sprat J., Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Bradshaw H.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U50311; AAA92313.1; -
DR HSSP: P56682; ICCV.
DR InterPro: IPR000561; -
DR InterPro: IPR002919; -
DR Pfam: PF01826; TIL; 2.
DR PROSITE: PS01186; EGF_2; 2.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 137 AA; 14948 MW; 35A3763760F88E30 CRC64;

```

Query Match 25.7%; Score 100.5; DB 5; Length 137;  
 Best Local Similarity 36.8%; Pred. No. 6.6e-05;  
 Matches 21; Conservative 7; Mismatches 26; Indels 3; Gaps 3;

QY 11 KCPSENRSCDRCQRCPCPNVVPKPLCIKICAPG-CVCRILGYLRNKKKVCVPRSKC 66  
 DB 81 KCPENETFEFCGTACEPTCEKPGPRP-CTRQCIYVNCQSSGFVRNGYR-CTELKEC 135

## RESULT 15

Q9U1U0 PRELIMINARY; PRT; 211 AA.  
 AC Q9U1U0; 01-MAY-2000 (TREMURrel. 13, Created)  
 DT 01-MAY-2000 (TREMURrel. 13, last sequence update)  
 DE 01-MAR-2001 (TREMURrel. 16, last annotation update)  
 DE Y69H2.10 PROTEIN.  
 GN Y69H2.10.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA McMurray A.A.;  
 RL Submitted (AUG-1997) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 investigating biology."  
 RL Science 282:2012-2018(1998).  
 DR EMBL; Z98877; CAB63407.1; -;  
 DR HSSP; P19398; IATB.  
 DR InterPro; IPR000561; -;  
 DR InterPro; IPR002919; -;  
 DR Pfam; PF01826; TIL; 3.  
 DR PROSITE; PS01186; EGF\_2; UNKNOWN\_2.  
 SQ SEQUENCE 211 AA; 22393 MM; 95D23AB933389A7 CRC64;

Query Match 25.7%; Score 100.5; DB 5; Length 211;  
 Best Local Similarity 37.3%; Pred. No. 9.6e-05;  
 Matches 22; Conservative 8; Mismatches 26; Indels 3; Gaps 3;

QY 8 GRGKCPSENRSCDRCQRCPCPNVVPKPLCIKICAPGCVCRILGYLRNKKKVCVPRSKC 66  
 DB 155 GTTKCSDEANWSKCH-NCKEYCFQTA-NPSC-KACWSGCCLDGFSKSTGTGLCVETAKC 210

Search completed: June 21, 2001, 10:53:22  
 Job time: 155 sec

